University of Warwick

Doctoral Thesis

Copula Methods in Econometrics

Author:

Kazim Azam

Supervisor:

Dr. Gianna Boero

Prof. Michael Clements

A thesis submitted in fulfilment of the requirements for the Doctorate of Philosophy

in the

Department of Economics

Contents

List of Figures

List of Tables

Acknowledgments

I am particularly grateful to Gianna Boero, Michael Clements and Michael Pitt for their continuous support and guidance, and to Valentina Corradi, Adam Johansen and Andrew Patton for their invaluable advices. I would also like to express my gratitude to Federica Liberini, Shereyar Malik, Nicola Pavanini and Craig Thamotheram. Comments from Michael Smith, Claudia Czado, Aleksey Min and Anastasios Panagiotelis at the 4th Workshop on Vine Copula Distributions and Applications (2011), and Copulae in Mathematical and Qunatitative Finance (2012) are gratefully acknowledged. I am also thankful for the valuable feedback from the participants of the the Warwick Lunchtime Workshop. Responsibility for the remaining errors is solely mine.

- Chapter 1 -

Introduction

This thesis consists of topics and issues related to copula modelling in Econometrics. Copula models provide an alternative to joint distribution analysis, which is frequently required in Economics and Finance.

Measures of dependence is mostly restricted to linear correlation among some random variables of interest. Embrechts et al. (2002) points out the limitations for such methods, as linear correlation is only one of the many measures of stochastic dependence. Multivariate Normal and t-distribution have frequently been used to measure dependence between assets, Hansen (1994), Harvey and Siddique (1999) and Engle and Manganelli (2004) employ such models for applications to risk-management and portfolio allocation. Other models such as Multivariate GARCH by Engle and Kroner (1995) and Dynamic Conditional Correlation (DCC) of Engle (2002) have also been applied, but they present various estimation problems in higher dimensions, and are also bounded by elliptical distributions for the multivariate analysis. In case a specific distribution is considered, problems can arise, however there are other methods to consider for multivariate analysis like Generalized Method of Moments (GMM) which require no assumptions on the distribution.

Recently Copula models have seen an increase in their use in finance and economics, though being around since Sklar (1959) theorem. They had success in applications, where dependence is non-linear and the random variables involved have different marginal distributions. It provides a framework which is general across different type of data types (marginal behaviour), unlike other joint non-linear modelling where problems are dealt with in a case-by-case method. There are various copula families available to capture complex dependence patterns, such as non-elliptical forms and tail dependence. In finance, Embrechts et al. (1999) show how for a Value at Risk (VaR) analysis, the assumption of multivariate Normal fails to capture joint observations in the tails, and hence apply copula methods. Cherubini and Luciano (2001) use them for pricing analysis of various assets. Embrechts et al. (2003) and Rodriguez (2007) study financial contagion through copula families. Bouyé et al. (2007) present a detailed coverage of copula methodology and other applications in finance. Common to all these works is the emphasis on how copula models provide a solution when we want to have joint analysis with either non-normal marginal distributions or mixture of different type of marginals. At the same time when joint distributions are no more best characterised by elliptical distributions, then using copula models avoids any possible misspecification of dependence measure through linear correlation.

In Economics instead, the copula based literature remains limited to few studies. Empirical analysis involving discrete data is unavoidable in economics, and alike using other joint modelling techniques, copula models represent complications. Munkin and Trivedi (1999), using discrete micro data show how generally joint modelling is troublesome, and the problem increases when the marginal distributions belong to different parametric families. Chib and Winkelmann (2001) specify a joint discrete distribution, without explicitly mentioning using a copula. Smith (2003) uses copula framework to study self-selection problem. Cameron et al. (2004) analyse a selection model with discrete outcomes in a copula framework. Demarta and McNeil (2005) among others, analyze categorical data from clinical trials. Zimmer and Trivedi (2006) employ a trivariate copula for dependency between health insurance status for married couples and their demand for health care. Trivedi and Zimmer (2006) provide details of their use in health economics applications. Patton (2006) introduces copulas in modelling of economic time series (continuous data). Hoff (2007) applies a Multivariate Gaussian copula on survey data of different types through Bayesian techniques.

We now give a formal definition to a copula function. According to Sklar (1959) theorem, any p-dimensional joint distribution H of some random variables $Y_1, \dots .Y_p$ can be decomposed to a copula C measuring their dependence, and their marginal distributions $F_1, \dots, F_p$, specifying their individual characteristics (fat tails, skewness etc.). Formally given as

$$H(y_1, \dots, y_p) = C(F_1(y_1), \dots F_p(y_p)).$$

Where $C : [0,1]^p \mapsto [0,1]$. The copula C is unique, if all the margins $F_1, \dots, F_p$ are continuous. The copula could also be stated as,

$$C(u_1, \dots, u_p) = \mathrm{pr}\ (U_1 \leq u_1, \dots, U_p \leq u_p).$$

u_j is the uniform variable computed through the marginal distribution, $u_j = F(y_j)$, where $j = 1, \dots, p$. The process of obtaining the uniforms is known as Probability Integral Transformation (PIT) (see Diebold et al. (1998)). If the joint distribution F is p-times differentiable, then by taking its p^{th} cross-partial derivatives we get

$$f(y_1, \dots, y_p) = \prod_{j=1}^{p} f_j(y_j) \cdot c(F_1(y_1), \dots, F_p(y_p)).$$

Such a decomposition provides a very flexible framework, where each F_j could belong to a different parametric family, and the dependence among the random variables is not confined to elliptical distributions (Gaussian or t-distribution). Nelson (2006) and Joe (1997) cover various statistical and mathematical properties of copulas, including estimation techniques.

We present the abstracts from the chapters of this thesis in the chronological order now.

Not only currencies are assets in investor's portfolio, but also central banks use them for implementing various economic policies. This can create some form of dependence among different exchange rates. We investigate the dependence pattern among the time series of daily Deutsch Mark (DM) (Euro later), Great Britain Pound (GBP) and the Japanese Yen (JPY) exchange rate, all considered against the U.S. Dollar during various economic conditions. To overcome the short-comings of marginal misspecification, and the restrictions of linear correlation, a flexible semi-parametric copula methodology is adopted where the marginals are non-parametric and the copula is parametric, to capture richer dependence form. Dependence is estimated as a constant measure and also allowed to vary over time. Our approach is the first time, where a time-varying copula parameter is considered in a semi-parametric setting avoiding any possible marginal misspecification, along with a depth full analysis of the dependence patterns among such vital currencies. During the Pre-Euro period, we find slightly more dependence when both DM (Euro)/USD and GBP/USD jointly appreciate as compared to joint depreciation, especially in the late 90s. Such results are reversed for GBP/USD and JPY/USD in the early 90s. In Post-Euro period, DM (Euro)/USD and GBP/USD exhibit stronger dependence when they jointly appreciate, which could indicate preference for price-stability in EU zone. Whereas the dependence of JPY/USD with both DM (Euro)/USD and GBP/USD is stronger when they jointly depreciate, this could imply preference for export competitiveness among the countries. In the beginning of Recent-Crisis period, DM (EURO)/USD and GBP/USD show stronger dependence when they jointly depreciate, but later during the period, we see the similar tendency for these exchange rates to be related more when they jointly appreciate. Such measures of asymmetric dependence among the exchange rates provide vital insight into Central banks preferences and investors portfolio balancing.

Multivariate analysis involving random variables of different data type like count, continuous or mixture of both is frequently required in econometrics. A Copula based methodology can be adopted for such data, where the association among the random variables is independently modelled from their specific marginal distributions. Depending upon the chosen marginal specifications, copula estimation proceeds. A semi-parametric copula estimation, where the marginals are specified empirically performs very well, but for discrete data its appropriateness is questioned (see Genest et al. (1995)). Hoff (2007) proposes a methodology where the marginal distributions are left completely unspecified and the copula parameters are estimated based on the order statistics of the observed data. We conduct an analysis to determine the effect on the estimates of a Gaussian copula due to various marginal specifications. The novelty of the work is that we are unifying all the bayesian approaches to copula estimation, to compare the effects of various marginal specifications on copula estimates, which is our contribution towards the literature on Markov chain based time series models. Through employing a Bayesian framework, we find that treating the marginal distributions as unknown outperforms both assuming an empirical distribution or misspecifying the marginal distributions, in terms of bias and mean square error for the estimates of the copula parameters. Hoff's method particularly outperforms the other specifications, when one or more of the marginals involved is of low count data type (binary).

Time series modelling can be very restrictive when accounting for various marginal specifications (non-normal distribution), data types and the dependence structure through time. On the contrary, Copula models allow such issues to be specified independently of each other. We propose a general technique to model a univariate strictly stationary time series through a copula. The novelty lies in the fact that it can be applied to both discrete and continuous data, and is invariant to any copula family. The technique is robust to any marginal mis-specification, and we successfully capture persistence in a time series through a copula. Expanding the methodology of Hoff (2007) for cross-sectional data, we set out a Bayesian sampling scheme to estimate the copula parameters, based only on the order statistics of the observed data. To show it's applicability, a real time series (weekly firearm homicides in Cape Town, South Africa) is used, and we are able to successfully capture the persistence in such a series. In terms of the Bayesian methodology, the technique performs well (fast mixing and low autocorrelation). Such a method provides flexibility in modelling non-gaussian based time series characteristics.

- Chapter 2 -

Dependence Analysis between Foreign Exchange Rates: A Semi-Parametric Copula Approach

§ 2.1 Introduction

Exchange rates are a vital aspect of International Economics, used to implement various economic policies. Along with GDP and interest rate, exchanges rates are an indicator of a country's economic outlook, and are determined through various cross-country economic fundamentals. In this era of globalisation, countries are not simply interested in closely monitoring their own currencies, but also the currency of other countries, which causes them to frequently intervene in the foreign exchange market. Such an intervention to guide their currency in a particular direction due to another currency, creates a dependence among the exchange rates. A synchronisation of business cycles, or difference in short-term interest rates across countries causing capital inflow/outflow, can also create comovement of exchange rates. Currencies are also held in investors portfolio along with other financial assets, and their preference over holding such currencies, also creates a relationship between the exchange rates.

Not only are the exchange rates dependent upon each other, they could also exhibit non-linear dependence and non-constant dependence through time. Takagi (1999) states if there are two countries, A and B, who export to foreign countries and in order to ensure their export prices are competitive to each other, then if country A's exchange rate depreciates, country B would ensure their exchange rate does too, which creates joint depreciation dependence. On the other hand if the countries prefer price stability among each other (maybe regional), then if country A's exchange rate appreciates, country B will intervene in the foreign exchange rate to ensure similar appreciation of their currency, and hence causing dependence due to joint appreciation. The variation in the preference of being competitive in terms of export, or ensuring price stability, creates an asymmetric dependence among the two currencies. Patton (2006) shows that Deutsch Mark and Japanese Yen (both against U.S. Dollar) before the introduction of the Euro, tend to exhibit stronger dependence when they jointly depreciate, as compared to when they jointly appreciate. Dias and Embrechts (2010) report similar results for the same currencies, but over different periods. Boero et al. (2011) show such asymmetric dependence patterns vary for different currencies. Another contributing reason for such asymmetric dependence patterns could be associated to the common denominating currency in the two exchange rates. U.S. Dollar has long been considered as a reserve currency, meaning investors prefer to hold it more in their portfolio as compared to the other currencies. So when U.S. Dollar appreciates, investors forgo their holdings of other currencies and shift their funds into the U.S. Dollar. On the other hand, when U.S. Dollar depreciates, they might not prefer to hold other currencies similarly. Such shifting of funds to and from the U.S. Dollar could also create an asymmetric dependence. These asymmetries are not only found in exchange rates, but also for other financial assets, Longin and Solnik (2001) show assets returns exhibit stronger dependence during market downturns as compared to market upturns. This paper aims to understand such phenomenon among Deutsch Mark (DM) (later Euro), Great Britain Pound (GBP) and Japanese Yen against the U.S. Dollar (USD) through different economic periods.

Patton (2006) adopts the copula methodology for time-series analysis and identifies higher dependence between DM (Euro) and JPY (both against dollar) when they are both depreciating, as compared to when they are both appreciating. Dias and Embrechts (2010) also report similar results over different sample period for both exchange rates.

Both Patton (2006) and Dias and Embrechts (2010) report such results for a constant and time-varying measure of dependence. Boero et al. (2011) perform similar analysis over several exchange exchange rates including GBP against the U.S. Dollar, but only for a constant measure of dependence over time. All these work show that dependence varies over different periods, for example before and after introduction of Euro. We extend previous analysis and investigate dependence between exchange rate before and after the euro, and over the recent financial crisis period.

Copula methodology requires the decomposition of the marginal distributions of the random variables from the dependence among them. Patton (2006) and Dias and Embrechts (2010), adopt a fully parametric approach, where the marginals are chosen from a parametric family along with a parametric copula. In their case, they assume the exchange rate returns to be specified through t-distribution, with varying degrees-of-freedom. The copula estimation relies on no misspecification of the marginals, and hence any parametric family chosen requires testing for the appropriateness of the chosen marginals specification. It is easy to misspecify the marginals, especially when the time-series in question is of high frequency (daily exchange rates). Genest et al. (1995), show how non-parametric marginals produce consistent and asymptotically normal copula estimates, given the unspecified margins are of continuous type. Kim et al. (2007) also show such a specification to produce efficient results for sample size larger than 100. Boero et al. (2011) adopt such an approach, which is generally termed as semi-parametric copula estimation. We extend the approach of Boero et al. (2011) to a time-varying dependence measure within a semi-parametric copula framework, which is the first time such a flexible approach to avoid any marginal specification has been considered in a dynamic setting.

We employ two copula families, the Gaussian copula and Symmetrized Joe-Clayton (SJC) copula of Patton (2006), to measure dependence patterns between DM (EURO)/USD, GBP/USD and JPY/USD. The time-varying measure of dependence for both copula evolves according to a ARMA type process, same as Patton (2006). The SJC copula is a two parameter based copula measuring the lower and upper joint tail dependence separately. The Gaussian copula acts as a benchmark specification. First, we filter the daily returns through a ARMA(p,q)-GARCH(1,1) model to obtain i.i.d observations required for the copula methodology, and then using the filtered returns we estimate both constant and time-varying measure of dependence between the 6-pairs for before (1990 - 1998) and after Euro (1999 - 2006), and for the recent crisis (2007 - 2009).

Using the Gaussian copula, we find strong dependence between DM (Euro)/USD and GBP/USD, especially before the introduction of Euro, which can be associated to how GBP/USD shadowed the DM (Euro)/USD in the early 90s. The dependence between DM (Euro)/USD and JPY/USD is much smaller compared to the previous pair, and similarly for GBP/USD and JPY/USD. The SJC copula however provides a better fit in terms of likelihood, and reports some asymmetric tail dependence patterns for GBP/USD and JPY/USD. After the introduction of the Euro, the SJC copula reports higher probability of joint appreciation between DM (EURO)/USD and GBP/USD, which could be associated with higher co-operation within the EU for price stability in the region. Both DM (Euro)/USD and GBP/USD when paired with JPY/USD show similar asymmetric results, but the time-varying measure reports periods where there is a higher probability to jointly depreciate as compared to probability of jointly appreciating. In the recent-crisis period, for DM (Euro)/USD and GBP/USD, the constant SJC copula measure could be misleading as both tend to jointly depreciate with greater probability, then compared to joint appreciation in the beginning of 2007, but later revert back to appreciating jointly with greater probability. This could be associated to the uncertainty the crisis caused in the beginning of the crisis, and both countries (being in the EU region) adopting an export competitive behaviour, whereas the dependence between JPY/USD and the other two exchange rate seems weaker. Generally our results indicate greater preference for price stability between DM(Euro)/USD and GBP/USD, which is understandable after the integration of the EU. Whereas, when paired with JPY/USD, more export competitive behaviour is suggested and investors view JPY/USD as an alternative to DM (Euro)/USD and GBP/USD, hence the negative correlation reported in time-varying Gaussian copula.

We start our analysis by first explaining the copula methodology in Section 2, where details over the constant and time-varying dependence measure is provided, along with non-parametric margins. Section 3 sets out the data, and describes some vital summary statistics. In Section 4, we present the results for both the constant and time-varying copula measures and present some economic intuition for the result. Finally concluding in Section 5.

§ 2.2 Semi-Parametric Copula Framework

We presented the copula definition for a multivariate case of dimension p in Chapter 1. In our empirical analysis, we are interested in capturing dependence among two random variables at a time, hence from now we will present the specifications for $p = 2$ (bivariate). Instead of denoting the random variables as Y_1 and Y_2, we denote them as X and Y and their respective PIT as u and v.

2.2.1 Copula Families

There exist a wide array of copulas families to chose from, depending upon the type of dependence a practitioner is interested in capturing. Nelson (2006) describes most of the commonly used copulas. Our main aim is to capture any asymmetric dependence among exchange rates, and show that in the presence of such asymmetries, a Gaussian copula would provide an inferior fit and fail to capture vital aspects. We will consider only two copula families for our analysis, the Gaussian copula and the Symmetric Joe-Clayton (SJC) copula. The latter is a two parameter based copula, one parameter capturing the joint upper tail dependence and the other parameter the joint lower tail dependence. In certain instances a one parameter copula like Clayton copula might provide a better fit, but we are also interested in the dynamics of asymmetric dependence through time, and see how both tail dependence measure evolve.

2.2.1.1 Gaussian Copula

A Gaussian copula is the most used copula, along with a t-copula from the elliptical set of distributions. It is analogous to a multivariate Normal distribution, when the margins are assumed/chosen to be normally distributed. It is a symmetric and zero tail dependence copula. For a bivariate case with uniform i.i.d random variables u and v between $[0,1]$ (obtained through marginal specifications), the Gaussian copula is given as

$$\begin{aligned} C_g(u,v|\rho) &= \Phi_g(\Phi^{-1}(u),\Phi^{-1}(v);\rho) \\ &= \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \\ &\times \left\{\frac{-(s^2-2\rho st+t^2)}{2(1-\rho^2)}\right\} dsdt, \end{aligned}$$

where Φ is the Cumulative Distribution Function (CDF) of a standard normal distribution, $\Phi_g(u,v)$ is a standard bivariate normal distribution and ρ the correlation parameter defined over $[-1,1]$.

2.2.1.2 SJC Copula

In order to capture asymmetric dependence in the tails, we have to employ a copula which separately parameterizes the left and the right tail. Joe (1997) proposes a copula termed as "BB7", also referred as Joe-Clayton copula. It is a two parameter based copula, given as

$$C_{JC}(u,v;\tau^U,\tau^L) = 1-(1-\{[1-(1-u)^\kappa]^{-\gamma}+[(1-v)^\kappa]^{-\gamma}-1\}^{-\frac{1}{\gamma}})^{\frac{1}{\kappa}},$$

where,

$$\begin{aligned} \kappa &= \frac{1}{\log_2(2-\tau^U)}, \\ \gamma &= -\frac{1}{\log_2(\tau^L)}. \end{aligned}$$

When $\kappa = 1$, Joe-Clayton copula reduces to Clayton copula, and when $\gamma \to 0$ it reduces to Joe Copula. τ^U and τ^L are the parameters of the Joe-Clayton copula, capturing tail dependence.

If the limit

$$\lim_{\delta\to 1}\Pr[U\le\delta|V\le\delta] = \lim_{\delta\to 1}\Pr[V\le\delta|U\le\delta] = \lim_{\delta\to 1}(1-2\delta+C(\delta,\delta))/(1-\delta) = \tau^U$$

exists, then the copula exhibits upper tail dependence if $\tau_U \in (0, 1]$ and no upper tail dependence if $\tau_U = 0$. In same manner, if the limit

$$\lim_{\epsilon \to 0} \Pr[U \le \epsilon | V \le \epsilon] = \lim_{\epsilon \to 1} \Pr[V \le \epsilon | U \le \epsilon] = \lim_{\epsilon \to 1} C(\epsilon, \epsilon)/\epsilon = \tau^L$$

exists, then the copula exhibits lower tail dependence if $\tau_L \in (0, 1]$ and no lower tail dependence if $\tau_L = 0$.

Patton (2006) points out that Joe-Clayton copula tends to report asymmetric dependence, even if the dependence in both tails is perfectly symmetric. He proposes a slight modification to the Joe-Clayton copula and terms it as "Symmetric Joe-Clayton" (SJC) copula. It is computed as

$$C_{SJC}(u, v;\ \tau^U, \tau^L) = 0.5(C_{JC}(u, v;\ \tau^U, \tau^L)\ +\ (C_{JC}(1-u, 1-v;\ \tau^U, \tau^L)\ +\ u\ +\ v\ -1)).$$

It treats symmetry as a special case and is consistent in reporting any asymmetry. An alternative technique would be to estimate an asymmetric measure of Kendall's tau for joint movements below zero (downwards) and then separately for joint movements above zero and compare those estimates with SJC copula, however for the SJC copula we are not able to derive the Kendall's tau equivalent of the joint upper and lower tail dependence parameters. By fact they are two different measures, SJC copula measures the dependence in the extreme of both tails, and an asymmetric Kendall's tau measure will simply be defined with a threshold at zero.

2.2.2 Time-Varying dependence

There is evidence that dependence among financial assets does not stay constant over time (see Bouyé et al. (2008) and Longin and Solnik (2001)). Such dynamics have great implications from portfolio diversification perspective and can identify how two assets behave jointly in various economic conditions. Given that we are dealing with exchange rates which tend to be highly volatile, we should account for changes in the contemporaneous dependence.

Similar to Patton (2006), we let the dependence parameter evolve according to an

ARMA process, both for the Gaussian and the SJC copula dependence parameters. We assume the functional form of the copula remains constant over time, but the copula parameters can evolve with time. As Patton mentions, the problem lies in defining the "Forcing Variable" for the evolution equation, as there is uncertainty to what causes the variation in the parameters. First we define the evolution of the upper and lower tail dependence parameter in the SJC copula. Identifying the terms for the evolution of such parameters is not easy in case of observation driven models. We adopt the same specification as Patton (2006) for both tails, given as

$$\tau_t^U = \Lambda\Big(\omega_U + \beta_U \tau_{t-1}^U + \alpha_U \cdot \frac{1}{10}\sum_{j=1}^{10} |u_{t-j} - v_{t-j}|\Big), \tag{2.2.1}$$

$$\tau_t^L = \Lambda\Big(\omega_L + \beta_L \tau_{t-1}^L + \alpha_L \cdot \frac{1}{10}\sum_{j=1}^{10} |u_{t-j} - v_{t-j}|\Big), \tag{2.2.2}$$

where $\Lambda(x) \equiv (1 + e^{-x})^{-1}$ is the logistic transformation, which keeps τ^U and τ^L bounded to $(0, 1)$. Both (2.2.1) and (2.2.2) are similar to an ARMA(1,10) process, where both the upper tail τ_t^U and the lower tail τ_t^L at period t depend upon their respective 1-period lag and a forcing variable for the time-varying limit probability, which is the mean absolute difference between u_t and v_t over the last 10 observations. Different specifications were also tried, but yielded no major improvements, so we adopted the dynamics specified as of Patton (2006).

The mean value in both (2.2.1) and (2.2.2) is inversely related to the concordance ordering of the copulas, value of zero corresponds to perfect positive dependence, 1/3 corresponds to independence and 1/2 implies perfect negative dependence. The choice of the forcing variable makes a good case, as under perfect positive dependence all u_t and v_t would be on the main diagonal of the copula support, and under independence scattered through out the support. For So the average distance from the point to the main diagonal is acts like an approximation for how close the last ten values of u and v (in time) were to being perfectly dependent, as difference of zero would imply perfect dependence and α will have a negative sign. This is indeed a dependence measure, which equates to the value being equated on the LHS in both equation (2.2.1) and (2.2.2). We can see that from

Figure 2.1. Another point to note is the forcing terms in both equations are bound by the range of value admissible by the LHS variable, as its the absolute difference between two uniform $[0, 1]$ values, hence the average difference will also be bounded between $[0, 1]$.

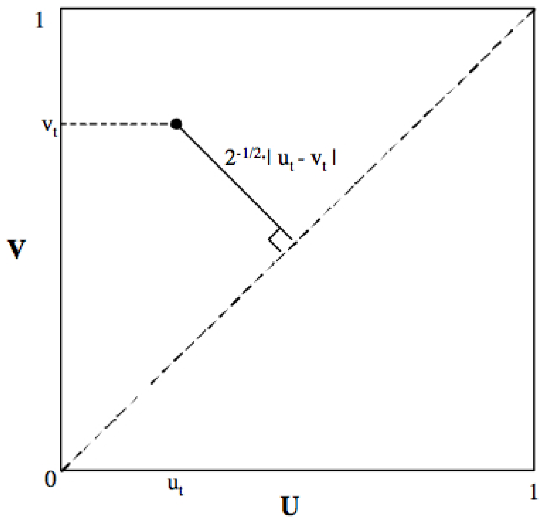


Figure 2.1: *Perfect dependence distance*

As we saw earlier the Gaussian copula only has one dependence parameter ρ, and we specify its evolution as

$$\rho_t = \tilde{\Lambda}\Big(\omega_\rho + \beta_\rho \cdot \rho_{t-1} + \alpha \cdot \frac{1}{10} \sum_{j=1}^{10} \Phi^{-1}(u_{t-j}) \cdot \Phi^{-1}(v_{t-j})\Big), \tag{2.2.3}$$

where $\tilde{\Lambda}(x) \equiv (1 - e^{-x})(1 + e^{-x})^{-1} = \tanh(x/2)$ is the modified logistic transformation required to keep ρ_t in [-1,1] at all instances. The evolution of ρ is similar to the one of SJC copula parameters, where in (2.2.3) the lag ρ_{t-1} captures any persistency in the dependence parameter. To be able to compare the SJC copula dynamics with the Gaussian copula dynamics a similar MA term is included, which is the mean of the product of the last 10 standard normals obtained through $\Phi^{-1}(u_{t-j})$ and $\Phi^{-1}(v_{t-j})$. Here this forcing term is equivalent to efficient Van der Waerden normal-scores rank correlation coefficient, which again equates to the correlation estimate on the LHS of the equation, being a correlation coefficient it is also bounded in between $[0, 1]$ to correspond by the values allowed by the LHS.

From the non-structural equations (2.2.1) - (2.2.3) we can easily compute the 1-step ahead forecast for the dependence measure, as for 1-step forecast we would have the observed x and y (returns in our case) and hence the corresponding u and v. This implies the forcing terms in all the equations can be computed. For the given set-up, we are unable to compute dynamic forecasts, for which we would require to change the forcing terms or be able to first forecast x and y (which would give us predicted u and v), maybe either through a parametric ARMA(p,q) - GARCH(1,1) set-up, or even a multivariate equation which would account for the covariance matrix.

2.2.3 Marginal Specification

We just stated the specifications related to the copula families to be used in this paper. Both the Gaussian and the SJC copula are parametrically specified. Before a copula is estimated, we need to compute u and v. Let n be the total number of observations, $i = 1, \ldots, n$, and F and G be the marginal distribution function for x and y respectively.

Copula modelling relies upon the assumption that the margins have i.i.d observations, but daily exchange rate returns tend to exhibit serial correlation and high volatility. Following the previous literature (see Patton (2006), Dias and Embrechts (2010) and Boero et al. (2011)), we first apply an ARMA(p,q)-GARCH(1,1) with normally distributed error term, on each exchange rate return series, through which we obtain the filtered residuals each assumed to be now independent and identically distributed. Such filtration still preserves the contemporaneous dependence among the returns. The order of p and q for the series along with the results are provided in Table 2.4.

After obtaining the filtered returns, we have to decide the functional form of F and G. In practice, the true marginal distribution function are not completely known, and if F and G are misspecified, the employed copula will also be misspecified. An assumed parametric family for each margin requires careful testing for any misspecification. Assuming that all the margins are continuous, we can adopt the approach of Genest et al. (1995), where the margins are left unspecified and computed non-parametrically based on the observed ranks. Boero et al. (2011) adopt the same approach for the filtered returns. So u and v are computed as

$$u = \widetilde{F}(x) = \tfrac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x),$$

$$v = \widetilde{G}(y) = \tfrac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(Y_i \leq y),$$

where $\mathbb{1}(.)$ is an indicator function and we divide the summation by $n+1$ to avoid CDF boundaries. $\widetilde{F}$ and $\widetilde{G}$ are employed for all $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ respectively. For complete clarity, the marginal distribution is not purely non-parametric for the returns, as first a parametric specification of ARMA (p,q)-GARCH (1,1) is applied to obtain filtered i.i.d returns. Only then from the filtered returns to obtain uniform random variables for input arguments to the copula a non-parametric (empirical CDF) specification is adopted. The novelty of our work lies in that it is the first time a non-parametric (apart from the filtering) approach has been considered to analyse the time-varying dependence.

Now we can proceed with the estimation of the copula parameters.

2.2.4 Estimation

Given we specified the copula parametrically and the margins non-parametrically, a semi-parametric estimation technique follows. Let Θ denote the parameter vector associated to a copula C, required to be estimated. The estimation is performed in two steps, first the pseudo observations $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are computed through $\widetilde{F}$ and $\widetilde{G}$, respectively. Then the second step involves maximum likelihood estimation of the pseudo log-likelihood function

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log\, c(\widetilde{F}(x_i), \widetilde{G}(y_i); \Theta),$$

where c denotes the copula density function. Genest et al. (1995) states the semi-parametric estimator $\widehat{\Theta}$ is consistent and asymptotically normal under suitable regularity conditions. Kim et al. (2007) show such an estimator to be robust when the margins are misspecified. Alternatives to the above estimator would be Inference Function for Margins (IFM) and Maximum Likelihood (ML) in fully parametric setting (see Joe (1997)). However in a fully parametric setting, the assumptions on the margins would have to be tested, for example Patton (2006) provides a goodness-of-fit test for the margins. Our approach avoids such issues due to the assumed margins, and only relies on the assumption of the random variables being continues. Kim et al. (2007) show the semi-parametric estimator to be as efficient as ML, when the sample size is larger than 100.

$\widehat{\Theta}$ is the estimated constant copula parameters. We are also interested in capturing any possible dynamics in the dependence parameters through an ARMA process, as described in Section 2.2.3. The parameters for both the Gaussian copula evolution $(\omega_\rho, \alpha_\rho, \beta_\rho)$ and for the SJC copula evolution $(\omega_U, \alpha_U, \beta_U, \omega_L, \alpha_L, \beta_L)$ are estimated through maximum likelihood. The estimated constant copula parameters act as the starting values for the time-varying dependence measure (i.e. $\widehat{\rho} = \rho_1, \widehat{\tau}^U = \tau_1^U, \widehat{\tau}^L = \tau_1^L$).

§ 2.3 Short-comings (Dynamic Copula)

Choosing and specifying the dynamics of a time-varying parameter are difficult, irrespective of it involving a copula model. For equations (2.2.1) - (2.2.3), we had to find an appropriate forcing term for the dependence parameter over time. Apart from the chosen forcing variable, we also tried other specifications, like weighting the u and v observations to how close they are to the extreme values, and using an indicator based on whether the observations were in the first, second, third or fourth quadrant. Such variations did not yield any improvements to the one we used. In all the equations (2.2.1) - (2.2.3), the values permissible by the terms entering the functional form have the same range as the dependent variable on the LHS, as they are a approximate measure of dependence through the previous periods. A drawback to our specification, is the less formidable forcing variable, but this is something commonly encountered in observation-driven based time series models. It is difficult to map out the true exogenous variation for the dependent variable of interest. Also previously stated, our approach restricts us for conducting dynamic forecasts, which is a strongly desired feature in time series literature. An alternative model to allow for time variation in the correlation is Dynamic Conditional Correlation (DCC) of Engle (2002), it offers great flexibility and ease to estimate the time-varying correlation.

It is in literature commonly estimated over two-steps, where first a GARCH is specified for each univariate series and then the equation for the correlation is estimated. We provide now a detailed simulation benchmark comparison of our copula specification against the DCC. The comparison can only be done with the time-varying Gaussian copula, as the DCC permits dependence to be measured only up to the level of correlation, so unlike copula specifications, no time-varying tail dependence or non-elliptical based time dependence can be estimated.

2.3.1 Simulation

We will generate 5 different time-varying correlation patterns, as given in Engle (2002), with an addition of a correlation pattern where the extremes of -0.99 and +0.99 are observed with high frequency.

- $\rho_t = 0.9$ (Constant)
- $\rho_t = 0.5 + 0.4\cos(2\pi t/200)$ (Sine)
- $\rho_t = \cos(t/4)$ (Fast Sine)
- $\rho_t = 0.9 - 0.5(t > 500)$ (Step)

T is fixed to 1000 (observations), and after each simulated ρ_t, we generate two corresponding random variables through separate Gaussian GARCH(1,1) processes given by

$$
\begin{aligned}
h_{1,t} &= 0.01 + 0.05r_{1,t-1}^2 + 0.94h_{1,t-1}, \\
r_{1,t} &= \sqrt{h_{1,t}}\epsilon_{1,t}, \\
h_{2,t} &= 0.5 + 0.2r_{2,t-1}^2 + 0.5h_{2,t-1}, \\
r_{2,t} &= \sqrt{h_{2,t}}\epsilon_{2,t}, \\
\rho_t &= E_{t-1}\epsilon_{1,t}\epsilon_{2,t},
\end{aligned}
$$

where the first series is highly persistent. Then we estimate the time-varying correlation through our GARCH-copula (Gaussian) and the DCC model, and compare the Mean Square Error (MSE). We sample from each correlation pattern 200 times. In Table 2.1

Table 2.1: *MSE from DCC and GARCH-Copula*

Correlation Type	DCC	GARCH-Copula
Constant	0.0023	0.0030
Sine	0.0292	0.0514
Fast Sine	0.0041	0.0067
Step	0.0066	0.0067

we report the MSE from both time-varying specifications. The MSE from DCC based time-varying correlation better produces the smaller error in all cases, even though in terms of the size, they are quite similar. The DCC specification follows the true correlation pattern quite well, whereas the GARCH-copula's forcing variable is smoothing over the last 10 observations and in doing so misses the rapid correlation changes. We also present a figure for the Fast Sine based correlation pattern of both the models for one of the Monte Carlo replication. From Figure 2.2 we see the plotted GARCH-copula fitted dynamics (green) does not follow the true correlation, whereas the DCC based correlation (red) does it better. As mentioned this is due to the smoothing of the forcing term, we could reduce the lags over which we take the expectation in such case. But as we reduce the lags, the fitted correlation becomes quite unstable. In terms of the modelling approach the dynamics in such a case are best given by a DCC model. The essence of using a GARCH-Copula would be to the case for where we might have non-normal joint distribution, where a non-elliptical distribution best fits the dependence structure. In such cases measures of dependence beyond correlation have to be adopted. For example, if two random variables exhibit greater dependence in lower values (lower tail dependence), then measures of correlation through a multivariate Normal would not appropriately estimate it, as it assumes no tail dependence (zero correlation within the tails), and we would have to seek some other methodology, like a Clayton copula for instance.

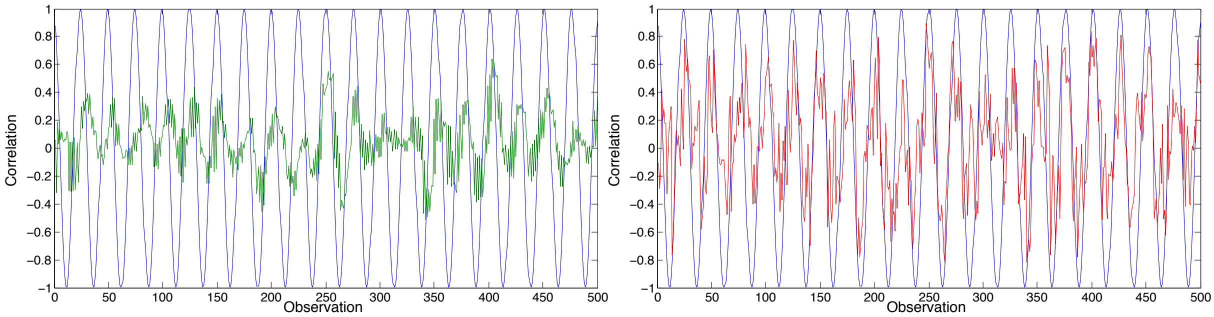


Figure 2.2: *Fitted Correlation: GARCH-Copula & DCC*

§ 2.4 Data

The data consists of daily exchange rates for Deutsch Mark (DM) (later converted at the conversion rate of Euro), Great British Pound (GBP) and Japanese Yen (JPY). All these currencies are denoted against the U.S. Dollar (USD). The full sample is over the period of 1st January 1990 up to 31st December 2009 and collected from Bank of England database[1]. We converted all the series to obtain log-differenced returns.

Three sub-samples are considered from the full sample. First, the Pre-Euro period from 1st January 1990 to 31st December 1998 (2276 observations). Second, the Post-Euro period from 1st January 1999 to 31st December 2006 (2020 observations) and finally the Recent-Crisis period from 1st January 2007 to 31st December 2009 (760 observations). We present time series plots for the three exchange rates in Figure 2.6 - 2.8, and summary statistics in Table 2.3 for the returns series. Time plots for DM(EURO)/USD and GBP/USD, show similar trends throughout the sample, especially in the Post-Euro period, when both currencies heavily appreciate together. This is also confirmed by the linear correlation values in Table 2.2, which shows strong correlation in both exchange rates, even in different sub-samples. JPY/USD on the other hand does not seem to follow any particular trends with DM(Euro)/USD or GBP/USD, and the correlation seems to be much weaker, becoming negative with GBP/USD in the Recent-Crisis period.

[1]http://www.bankofengland.co.uk

Table 2.2: *Pair-Wise Linear Correlation*

	Pre-Euro			Pre-Euro			Recent-Crisis		
	EURO	GBP	JPY	EURO	GBP	JPY	EURO	GBP	JPY
EURO	1			1			1		
GBP	0.719	1		0.633	1		0.651	1	
JPY	0.500	0.352	1	0.343	0.355	1	0.124	-0.120	1

Table 2.3, shows all of the series have skewness and excess kurtosis. DM/USD in the Pre-Euro period has almost zero skewness and GBP/USD in the Post-Euro period has kurtosis of almost 3, but apart from these two cases, none of the other series can be described through a normal distribution. The Jarque-Bera Statistic rejects normality with very large values. The ARCH-LM test, suggests the presence of heteroscedasticity for most of the series, hence it is appropriate for us to employ an ARMA(p,q)-GARCH(1,1) type filtering for the returns series.

§ 2.5 Copula Results & Economic Interpretation

In this Section we present the results from both the constant and time-varying measure of dependence computed through the Gaussian and the SJC copula. We discuss the results in detail over the various sub-samples. We seek to answer few questions, first, whether dependence can be assumed to stay constant not simply across different economic conditions (over sub-samples), but also within a specific period (within a sub-sample). Secondly, whether there exist any particular asymmetric dependencies, and the possible reasons for such patterns.

2.5.1 Pre-Euro

Correlation measures over this period seem to be very high, as compared to the other sub-samples. The constant Gaussian copula reports correlation of 0.71 between DM(Euro)/USD and GBP/USD in Table 2.5, which is of course similar to the pair-wise linear correlation in Table 2.2. Such strong correlation is not surprising, as the Pound shadowed closely the Deutsch Mark since 1988 to tackle inflation. The time-varying Gaussian copula in Figure 2.9 suggests the dependence stayed quite constant among the pair, except by the end of 1996, where there is a slight decline to about 0.4. Such a decline could be due to the interest rate lowering announcement in August 1996 by Bank of England to tackle inflation. Lower interest rate causes investors to shift their funds from GBP (causing depreciation), but DM (Euro) did not get necessarily effected by it. The constant SJC copula results in Table 2.5 suggest no asymmetric dependence, as the differenced tail dependence measure (τ^U - τ^L) is insignificant at 5%. Although the time-varying tailed differenced series in Figure 2.3 shows after 1993 the difference in upper tail and lower tail to be negative, this could correspond to greater preference for price stability in the region. Overall, for dependence between DM(Euro)/USD and GBP/USD, the time-varying results show that dependence does not stay constant over this period, both for Gaussian and SJC copula measure.

The relationship between DM(Euro)/USD and JPY/USD seems very stable through this period. The constant Gaussian copula reports correlation of 0.52 in Table 2.5 and the time-varying Gaussian shows no deviation from this level in Figure 2.9. Such patterns, might be suggestive of the fact that these two countries shared similar economic conditions and had similar foreign trade patterns, which created a unique and constant tie between them. The constant SJC copula measure reports no asymmetric dependence in Table 2.5, but from Figure 2.3 we see the difference in the tails of about 0.1. The results indicate the correlation patterns through a Gaussian copula can be appropriately described by a constant measure (similar likelihood in Table 2.5 and Table 2.6), the time-varying SJC copula also does not reveal more information about the dependence through this period, than what the constant SJC copula reports.

GBP/USD and JPY/USD are among the most volatile currencies, and due to this volatility investors seek to gain profits from short buying and selling. The correlation is relatively lower compared to the previous pairs above, of 0.37. The constant Gaussian copula predicts the correlation fairly well untill 1996, where the correlation drops and reaches the minimum of 0.1. The constant measure of SJC copula in Table 2.5 suggests the difference between joint upper and lower tails to be -0.1 and significant at 5%, but from Figure 2.3, we see the difference in the tails is very volatile and changes sign frequently. To associate such changes due to some form of economic policy of one or both of the central banks would not be suitable. Investors hold various currencies in their portfolios and take positions which could imply they shift out (joint depreciation) of the two currencies in a similar manner. GBP and JPY are not considered as candidates for being a reserve currency, and investors frequently buy and sell them. Therefore a time-varying copula should be employed in order to provide a more adequate representation of the dependence between these pair of currencies.

Unlike Patton (2006), we report the dependence between DM (Euro)/USD and JPY/USD to be symmetric, but similar to Patton, the time-varying measure of differenced tail dependence is not zero. To remind again we follow a semi-parametric copula estimation, whereas Patton (2006) sets out a fully parametric copula approach and have a slightly larger backdating period.

2.5.2 Post-Euro

After the Euro was introduced, now DM (Euro) did not simply represent Germany, but some major European economies. Not only a single currency was introduced but the EU was strengthened, where trade policies among all European countries (including Great Britain) where agreed. The constant Gaussian copula again shows strong correlation of 0.64 between DM (Euro)/USD and GBP/USD, and the time-varying measure in Figure 2.10 suggests also such correlation to have stayed constant, except a dramatic fall in early 2001, which could be due to pessimism about the newly formed currency causing smaller proportion of DM (Euro) to be held in investors portfolio. The constant SJC copula measure reports a stronger tendency ($\tau_U = 0.36$) towards large joint appreciations ($\tau_L = 0.53$) with respect to USD, than towards joint depreciation. Such a result could be due to the strong bounds created by the EU and the preference for price stability through the EU, rather than export competitiveness. Although from the differenced time-varying SJC copula in Figure 2.4, we see that at the beginning of the period the difference sometimes in the tails is reversed, later on the lower tail dependence exceeds that the upper tail. Even though the constant SJC copula accurately predicts the directions of the asymmetry, it under predicts the magnitude which at points reaches up to -0.4. This asserts the point even strongly that among a unified EU pricing stability is more preferred as compared to having a preference for being competitive in exports. Also Euro is the currency for most of the European countries, and hence for UK to be competing the rest of the Europe is very unlikely. The constant SJC copula does report the right sign on the tail difference in the later half of this period, but the magnitude is surely not appropriate to represent the period.

The constant Gaussian copula no longer adequately captures the correlation pattern among the DM(Euro)/USD and JPY/USD. Figure 2.10, shows the correlation goes to negative values in the infancy of Euro. This again could be due to the uncertainty over the newly created currency, and investors regarding Yen as a more secure holding in their portfolios, as compared to Euro. Constant SJC copula measure indicates the tails to be symmetric, but the time-varying SJC copula in Figure 2.4 shows instances of upper tail dependence being greater than lower tail dependence, which is understandable as Japan is not really part of EU trade treaties and now an export competitive position is preferred. Within this period, for DM (Euro)/USD and JPY/USD, constant measure of dependence would be misleading and a time-varying copula would be more appropriate.

The correlation between GBP/USD and JPY/USD seems more stable and constant, also the time-varying Figure 2.10 does not show much deviation from the constant level. The constant SJC reports no asymmetry, but this is true for the beginning of the period, but later in the period as we see from Figure 2.4, there is a greater probability of joint depreciation as compared joint appreciation. The linear correlation for this pair of currencies can be specified through a constant Gaussian copula, but for asymmetric dependence the constant SJC copula fails to capture the variation in the joint tails.

We cannot compare the results here with previous literature, as our sample for post Euro is much longer and unlike other work the correlation/dependence attains stable values after the uncertainty due to the new currency.

2.5.3 Recent-Crisis

This period represents turmoil and uncertainty from many aspects. Investor do not know what currencies to hold. The crisis originated from the U.S. soon had spillover effects into major currencies. From the constant Gaussian copula results, we see the correlation between DM (Euro)/USD and GBP was almost the same as in previous periods. The time-varying measure reveals similar constant correlation until the end of 2008 when correlation dropped significantly, this could be associated to bail-outs of the UK banks. The SJC constant copula reveals again a significant (at 5%) asymmetry in the tail, where there is higher probability for these currencies to depreciate together. The time-varying SJC copula shows in Figure 2.5 that at the beginning of the crisis there is higher probability to depreciate together. The U.S. Dollar appreciated in the beginning of the crisis, which is quite unusual given the crisis originated from there. This was due to short-term interest rate differentials, which investors tried to take advantage of and hence moved away from Euro and Pound. But such directions were reversed as soon as the risk aversion abated. Through such times price stability in the EU was strongly among the agenda, and therefor we see a much stronger probability of joint appreciation between DM (Euro)/USD and GBP/USD. The time-varying measure for both copulas is more suitable for this pair of exchange rates in this period.

Between DM (Euro)/USD and JPY/USD, the correlation fell to 0.12. The time-varying Gaussian copula confirms this in Figure 2.11. By the mid 2008, the correlation becomes very volatile, which could be due to investors trying to seek safe portfolio holdings. The constant and time-varying SJC copula indicates no asymmetries in the tails.

The correlation between GBP/USD and JPY/USD became negative, -0.12. This is also confirmed in the time-varying Gaussian copula case, the correlation patterns in late 2008 is similar to the correlation between DM (Euro)/USD and JPY/USD, indicating similar positions for Euro and Pound as compared to Yen. The constant and time-varying SJC copula are not reported, due to zero tail dependence found. Constant copula fails to address the extent of negative correlation in late 2008.

In terms of the best copula specification, we see the time-varying SJC copula has the highest log-likelihood value. As all sub-samples are large, the Akaike Information Criteria reports the same best fitting copula.

Overall, we discussed few reasons for observing dependence patterns for the currencies considered, though there could be many more reasons for observing these patterns. We have not discussed the role of USD, which through out the years has served investors as a reserve currency and movements to/from USD to other currencies might not be the same. Exchange rate is not only an economic tool for policy implementation, they are also considered an asset along with other stock assets. But unlike other financial assets, investors hold projections over economic conditions which lead them to hold specific holdings on currencies, and this could create complex dependence patterns. We need to use the time-varying measure to have a full understanding of the dependence.

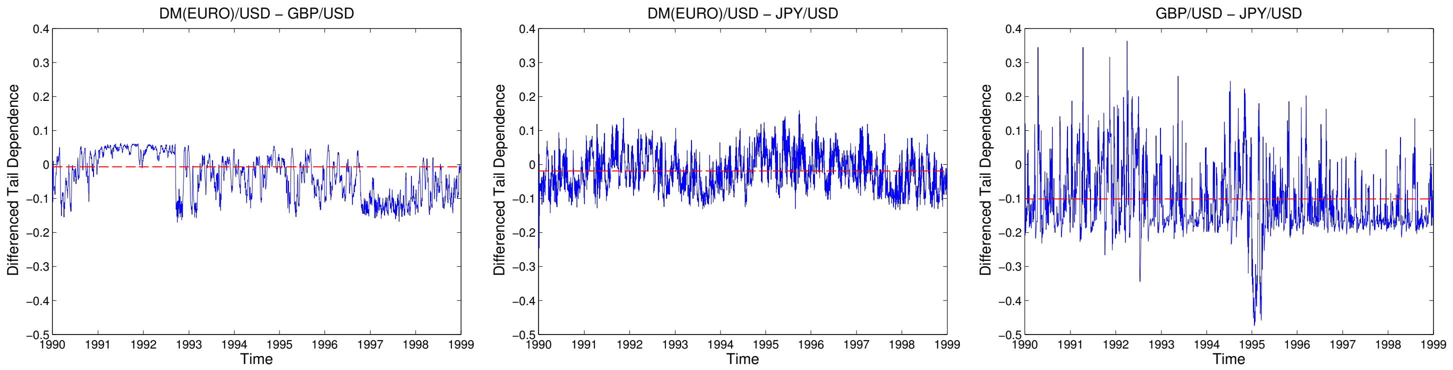


Figure 2.3: *Pre-Euro TV-SJC* $(\tau_t^U - \tau_t^L)$ *tail differences*

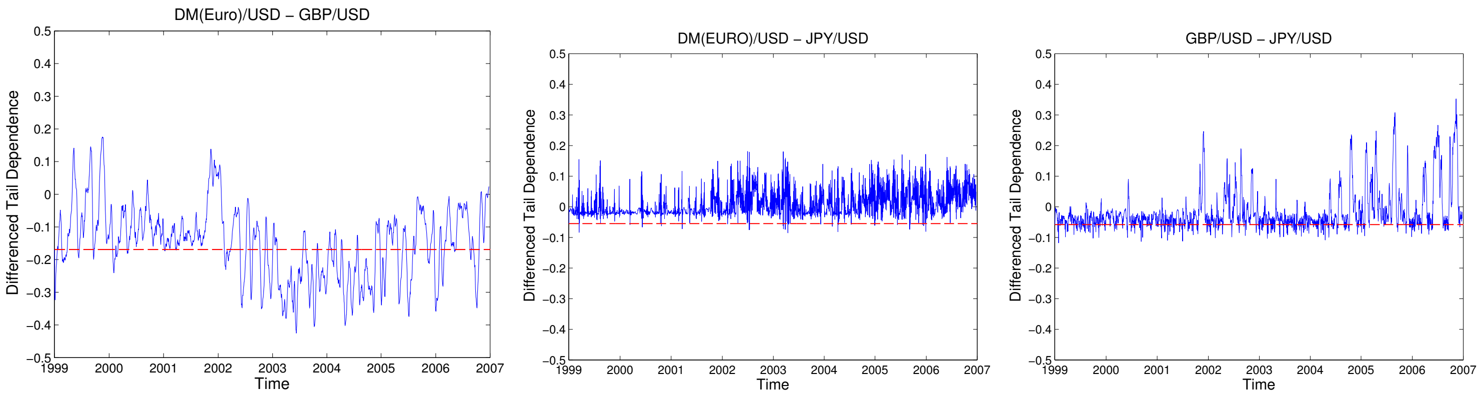


Figure 2.4: *Post-Euro TV-SJC* $(\tau_t^U - \tau_t^L)$ *tail differences*

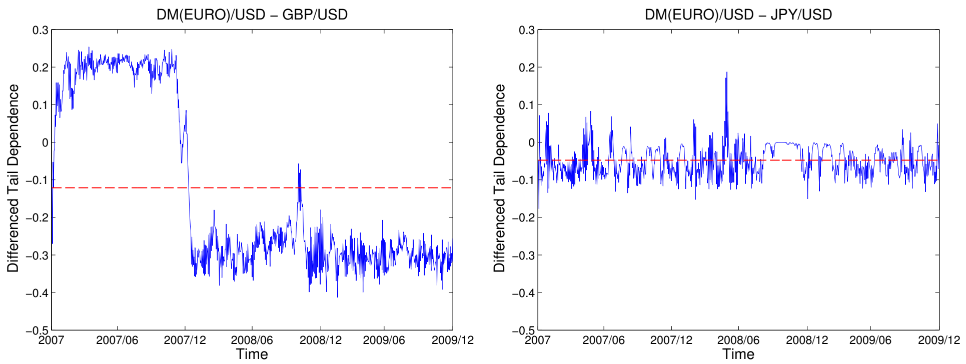


Figure 2.5: *Recent-Crisis TV-SJC* $(\tau_t^U - \tau_t^L)$ *tail differences*

DM (EURO)/USD and GBP/USD act very similarly and are driven in economic conditions regulated by the EU, this creates strong dependence and the European Bank and Bank of England to co-operate together towards price stability, and hence we observe strong dependence when they appreciate together. DM (Euro)/USD and JPY/USD prior to the Euro, show stable correlation, but after the introduction of the Euro the dependence is stronger when they jointly depreciate which could indicate a preference to stay competitive in terms of export prices. The relationship between GBP/USD and JPY/USD is quite volatile, as investors seek profitable holdings on these currencies. After the introduction of the Euro, GBP/USD a follows similar correlation with JPY/USD to that of DM (Euro)/USD and JPY/USD. We see constant measure of correlation/dependence do not reveal full information, there are times when correlation changes signs and magnitude, and hence we should employ the time-varying measures of dependence, as compared to assuming constant dependence.

§ 2.6 Conclusion

Various currencies are related to each other due to economic interaction among countries and how they are held in investor's portfolio. Their relationship in various economic conditions not only can reveal vital information to policy makers, but can also provide insight to investors for diversification purposes.

Given non-normality of daily exchange rates and joint non-linear dependence among exchange rate returns, we adopt a semi-parametric copula approach which overcomes the short-comings of multivariate Normal and t-distribution. Our approach is similar to Patton (2006) and Dias and Embrechts (2010), but unlike them we do not assume any parametric distribution for the marginals. Along with a parametric copula we specify the marginals to be non-parametric. Such a specification is robust to any misspecification of the marginals. Genest et al. (1995) show an estimator based on the ranks of the observed data is efficient and asymptotically normal for continuous data. Kim et al. (2007) report such a specification is robust to any misspecification of the marginals. Boero et al. (2011) employ a similar technique, but to estimate constant dependence only. We extended their approach to study dependence in a time-varying case.

We examine the dependence pattern between DM (Euro)/USD, GBP/USD and JPY/USD in different economic conditions. From using the Gaussian copula and the SJC copula, we see varying patterns of dependence in period before introduction of Euro, after and the most recent financial crisis.

We show linear correlation measures do not reveal dependence completely and to capture any possible asymmetric tail dependence we should adopt a two parameter copula like the SJC copula. In the Pre-Euro period DM (Euro)/USD and GBP/USD are highly correlated and such correlation persists through the other sub-samples. A time-varying analysis however shows that there are periods when the correlation weakens. From measuring asymmetric tail dependence, we find that the constant SJC copula fails to capture the variation in the joint tails, as there seems to be some pairs which have a higher probability to jointly appreciate as compared to probability of joint depreciation during different sub-samples. For DM (Euro)/USD and JPY/USD, correlation is quite constant as confirmed by the time-varying measure. There does not seem to be any particular preference from central banks to create export competitive environment or create price stability. The relationship between GBP/USD and JPY/USD seems very volatile through all the samples, and there is asymmetric tail dependence which the constant SJC copula does not completely capture. After Euro's introduction, DM (Euro)/USD and GBP/USD become more dependent when they jointly appreciate, reflecting the preference for price stability of both central banks, this is understandable as EU has trade policies in place, which are very co-operative and protect EU countries. Although there are certain periods (early Recent-Crisis period), where the probability to jointly depreciate is higher than probability to jointly appreciate, this could be due to shifting of funds into USD from both currencies. Both DM (Euro)/USD and GBP/USD have a similar stance towards JPY/USD, and hence the correlation between DM (Euro)/USD and JPY/USD and GBP/USD and JPY/USD show similar patterns.

We show how dependence evolves over time and assuming a constant dependence measure fails to capture the variations. The whole analysis is performed in a setting which ensure no misspecification of the marginal behaviours (distributions). We also prove how dependence patterns change with different economic conditions.

Table 2.3: *Sample Statistics*

	Pre-Euro			Post-Euro			Recent-Crisis		
	DM	GBP	JPY	EURO	GBP	JPY	EURO	GBP	JPY
Mean	−0.001	−0.001	−0.011	−0.005	−0.008	0.003	−0.011	0.025	−0.033
Median	0.0006	−0.013	0.019	0.000	−0.014	0.017	−0.029	0.020	0.010
Std.Deviation	0.666	0.607	0.736	0.620	0.510	0.622	0.727	0.814	0.811
Skewness	0.055	0.239	−0.787	−0.126	−0.083	−0.253	−0.257	0.073	−0.730
Kurtosis	5.244	6.328	9.147	4.391	3.748	4.546	8.253	7.330	6.811
Jarque-Bera Statistic	478.6	1007	3819	168.3	49.34	222.6	882.3	594.3	527.4
Arch-LM Statistic	77.29	114.8	110.4	4.572	7.377	7.706	74.00	43.60	34.04
Number of Observations	2276	2276	2276	2020	2020	2020	760	760	760

Note: ARCH-LM Test employed for presence of heteroscedasticity at 5 lags.

Table 2.4: *ARMA(p,q)-GARCH(1,1)*

	Pre-Euro			Post-Euro			Recent-Crisis		
	DM	GBP	JPY	EURO	GBP	JPY	EURO	GBP	JPY
C	-0.002	-0.004	-0.012	-0.012	0.004	-0.012	-0.033^{*}	-0.024	-0.004
	(0.013)	(0.014)	(0.011)	(0.013)	(0.013)	(0.011)	(0.019)	(0.027)	0.020
AR(1)	–	–	0.074^{***}	–	–	–	–	–	–
			(0.021)						
AR(3)	–	-0.040^{*}	–	–	–	–	–	–	–
		(0.022)							
GARCH Constant	0.007^{***}	0.011^{***}	0.002^{**}	0.003	0.009^{***}	0.005^{**}	0.001^{**}	0.006^{**}	0.002^{**}
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)	(0.000)	(0.003)	(0.001)
ARCH Term	0.044^{***}	0.067^{***}	0.040^{***}	0.020^{***}	0.026^{***}	0.032^{***}	0.053^{***}	0.069^{***}	0.062^{***}
	(0.006)	(0.007)	(0.004)	(0.005)	(0.006)	(0.008)	(0.009)	(0.013)	(0.012)
GARCH Term	0.948^{***}	0.914^{***}	0.954^{***}	0.972^{***}	0.950^{***}	0.947^{***}	0.947^{***}	0.927^{***}	0.937^{***}
	(0.008)	(0.008)	(0.005)	(0.008)	(0.012)	(0.013)	(0.006)	(0.011)	(0.011)

*** - 1%, ** - 5% and * - 10% significance level. GBP/USD and JPY/USD for the Pre-Euro period required the 1st and the 3rd lag, respectively.

Table 2.5: *Constant Copula Results for DM-GBP, EURO-JPY and GBP-JPY*

		Pre-Euro			Post-Euro			Recent-Crisis		
		DM-GBP	DM-JPY	GBP-JPY	EURO-GBP	EURO-JPY	GBP-JPY	EURO-GBP	EURO-JPY	GBP-JPY
Gaussian Copula	(ρ)	0.717***	0.516***	0.377***	0.635***	0.360***	0.355***	0.635***	0.116***	−0.103***
		(0.007)	(0.013)	(0.017)	(0.011)	(0.018)	(0.018)	(0.017)	(0.036)	(0.036)
	LL	−817.0	−349.5	−172.3	−520.4	−140.0	−142.31	−196.5	−5.168	−4.013
SJC	(τ^U)	0.547***	0.296***	0.133***	0.360***	0.161***	0.160***	0.404***	–	–
		(0.016)	(0.025)	(0.030)	(0.028)	(0.032)	(0.031)	(0.039)		
	(τ^L)	0.554***	0.315***	0.231***	0.530***	0.215***	0.218***	0.526***	0.064	–
		(0.015)	(0.025)	(0.026)	(0.016)	(0.029)	(0.029)	(0.026)	(0.045)	
	$(\tau^U_t - \tau^L_t)$	−0.007	−0.019	−0.101***	−0.169***	−0.005	−0.058*	−0.121***	−0.064	–
		(0.019)	(0.029)	(0.032)	(0.027)	(0.034)	(0.034)	(0.041)	(0.045)	
	LL	889.3	338.3	167.4	560.3	147.2	152.3	221.8	11.73	–

*** - 1%, ** - 5% and * - 10% significance level.

Table 2.6: *Time-Varying Copula Results*

		Pre-Euro			Post-Euro			Recent-Crisis		
		DM-GBP	DM-JPY	GBP-JPY	EURO-GBP	EURO-JPY	GBP-JPY	EURO-GBP	EURO-JPY	GBP-JPY
Normal										
	Constant	-0.342^{***}	0.645	-0.020^{***}	0.020^{***}	0.015^{*}	0.001	0.335^{*}	0.396	-0.225
		(0.000)	(0.800)	(0.001)	(0.007)	(0.008)	(0.006)	(0.194)	(0.136)	(0.154)
	α	0.107^{***}	-0.054	0.018^{***}	0.282^{***}	0.163^{***}	0.081^{***}	-0.158^{***}	0.762	0.557^{***}
		(0.000)	(0.066)	(0.001)	(0.031)	(0.028)	(0.020)	(0.052)	(0.153)	(0.169)
	β	2.857^{***}	1.030	2.123^{***}	2.114^{***}	1.962^{***}	2.052^{***}	2.095^{***}	-1.763	-1.875^{***}
		(0.012)	(1.487)	(0.008)	(0.025)	(0.041)	(0.033)	(0.239)	(0.186)	(0.167)
	LL	827.0	350.1	179.0	583.4	183.0	108.5	213.4	17.94	9.470
SJC										
	Constant(U)	1.130^{***}	1.505	3.650^{***}	-1.526^{***}	4.658^{***}	-0.253	-1.943^{***}	3.956	–
		(0.145)	(1.275)	(0.905)	(0.421)	(0.703)	(0.497)	(0.041)	(2.601)	
	$\alpha(U)$	0.911^{***}	-1.255	-3.014^{***}	3.684^{***}	-3.149^{***}	2.615^{***}	4.365^{***}	2.249	–
		(0.150)	(2.022)	(0.709)	(0.511)	(0.933)	(0.931)	(0.082)	(3.750)	
	$\beta(U)$	-9.471^{***}	-9.158^{***}	-22.65^{***}	-2.284^{*}	22.62^{***}	-7.637^{***}	-1.563^{***}	-23.13	–
		(0.127)	(3.667)	(4.659)	(1.303)	(2.762)	(2.029)	(0.372)	(9.380)	
	Constant(L)	1.984^{***}	2.059^{***}	-2.009^{***}	-1.993^{***}	2.355^{***}	1.256	2.842^{***}	3.919	–
		(0.062)	(0.496)	(0.032)	(0.012)	(1.167)	(1.289)	(0.629)	(0.730)	
	$\alpha(L)$	-0.636^{***}	-4.192^{***}	4.300^{***}	4.091^{***}	-1.307	-1.310	-3.451^{***}	-3.802	–
		(0.139)	(0.278)	(0.052)	(0.076)	(1.549)	(2.265)	(0.952)	(0.709)	
	$\beta(L)$	-8.582^{***}	-6.673^{***}	-0.761	-0.283^{***}	-14.52^{***}	-9.561^{***}	-5.061^{**}	-18.28^{***}	–
		(0.675)	(2.344)	(0.192)	(0.017)	(4.022)	(3.642)	(2.534)	(3.024)	
	LL	1024	355.60	189.4	668.3	224.3	188.1	251.67	34.00	–

*** - 1%, ** - 5% and * - 10% significance level.

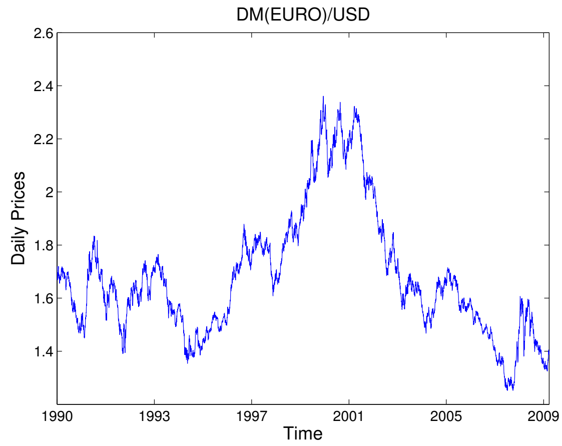


Figure 2.6: *Daily DM/USD Exchange Rate*

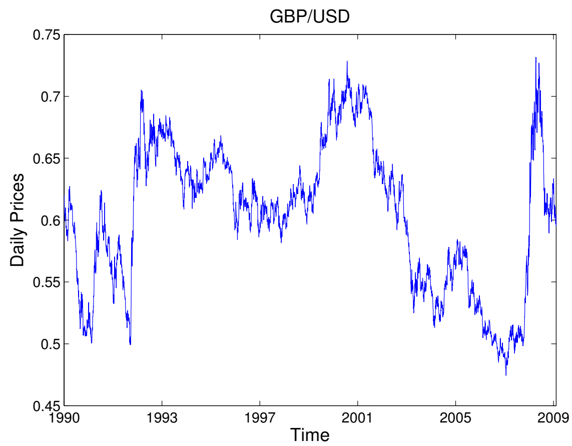


Figure 2.7: *Daily GBP/USD Exchange Rate*

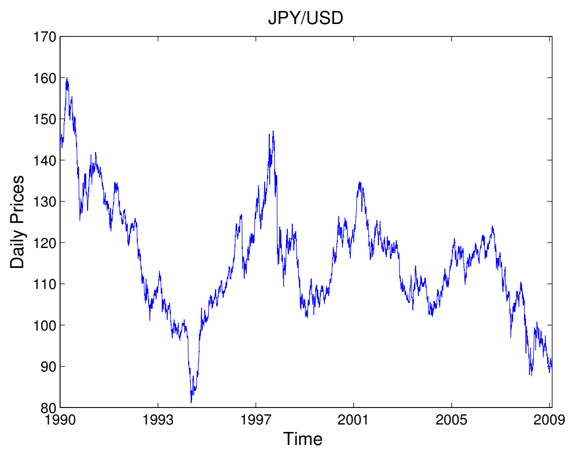


Figure 2.8: *Daily JPY/USD Exchange Rate*

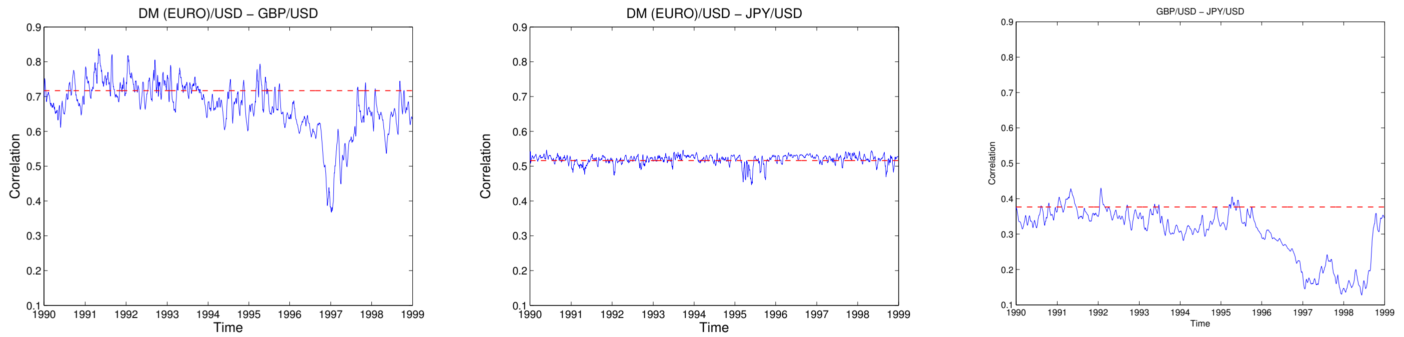


Figure 2.9: *Pre-Euro TV Gaussian Copula*

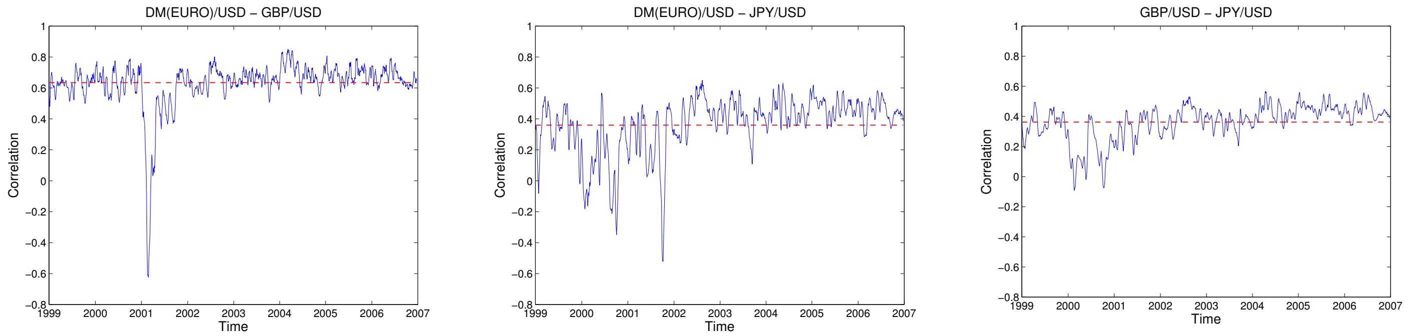


Figure 2.10: *Post-Euro TV Gaussian Copula*

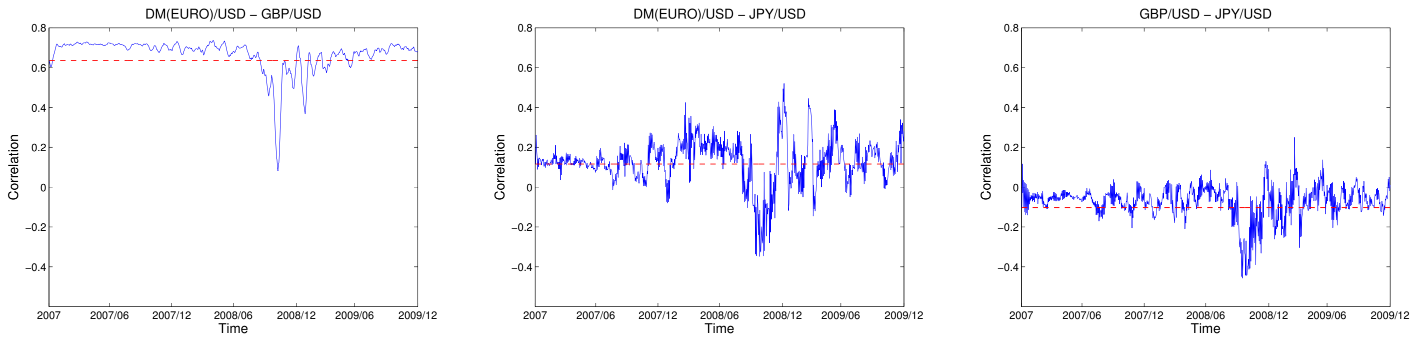


Figure 2.11: *Recent-Crisis TV Gaussian Copula*

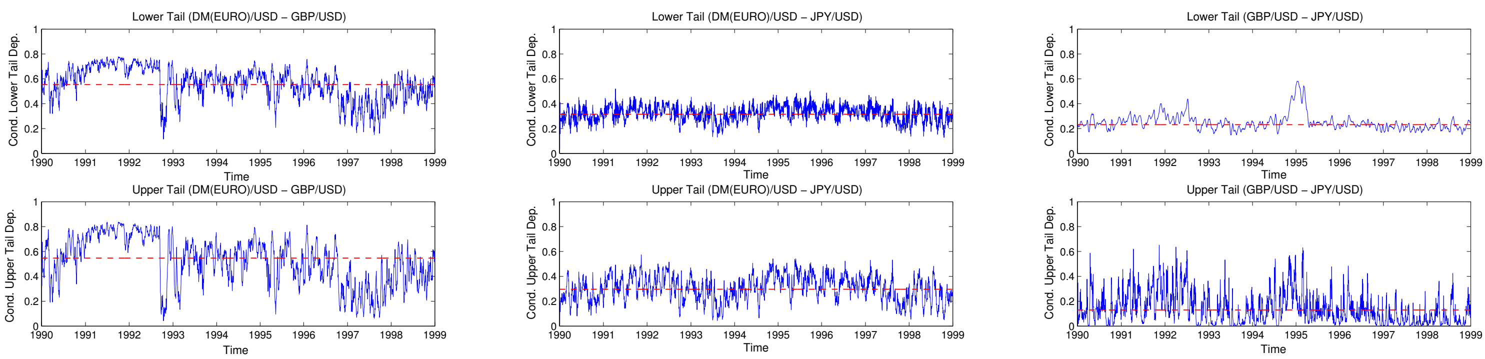


Figure 2.12: *Pre-Euro TV SJC Copula*

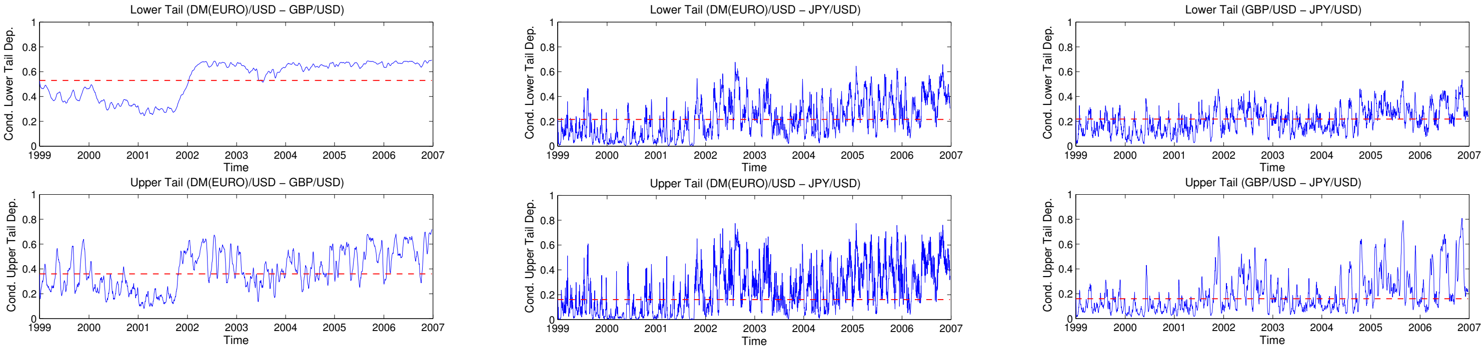


Figure 2.13: *Post-Euro TV SJC Copula*

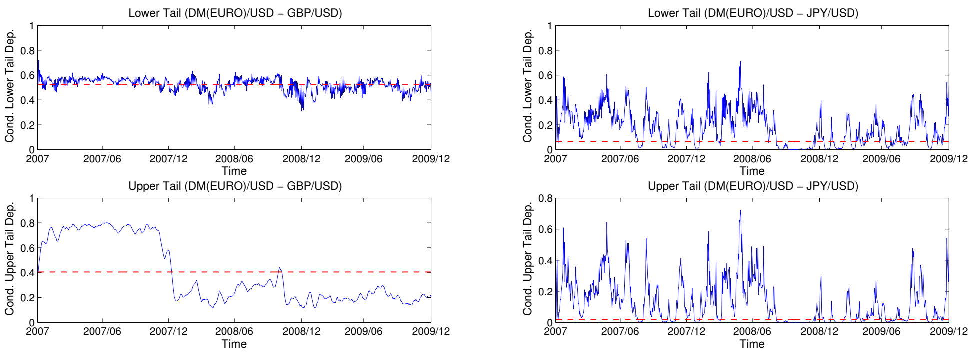


Figure 2.14: *Recent-Crisis TV SJC Copula*

- Chapter 3 -

Marginal Specifications and a Gaussian Copula Estimation

§ 3.1 Introduction

The advantages and flexibility of copula functions is most entertained under the assumption that the marginal distribution of the random variables is of continuous type. In case the data is of discrete outcomes, there are various difficulties faced, regardless of employing parametric or non-parametric marginal distributions. From a theoretical point, if the marginals involved are of discrete type then the copula function is not unique. Trivedi and Zimmer (2006) state that for discrete margins, the copula maximization often runs into computational problems, like failure of the algorithm to converge. They propose to employ a continuation transformation to the discrete variable and then base the likelihood estimation on continuous copula families, rather than working with differences of the copula probability mass functions. Genest and Nešlehová (2007) show that to use the rank-based estimators we first have to deal with the ties observed in the ranks (splitting, ignoring etc.), and such estimators tend be highly biased. Pitt et al. (2006) propose a Bayesian sampling scheme for continuous and discrete margins in a fully parametric Gaussian copula framework where some of the issues regarding discrete margins are dealt with, but some data (survey data) do not belong to standard parametric families.

Alternatively for discrete or mixture of continuous-discrete data, Hoff (2007) proposes a method where the marginals are left unspecified, and the copula estimation is based on the order statistics of the observed data using Bayesian techniques. The inference on the copula parameters is based on a summary statistic which is not a function of the nuisance marginal parameters.

In this paper we set out a simulation to study the effects on the copula parameters estimates, when we have mixed (continuous and discrete) type margins. We will employ various methods to deal with such margins. One such methods is to let the margins be empirically distributed, and the second is to perform a continuous transformation to the discrete margins and hence mispecify the margins. Our aim is to compare these two specifications against Hoff's method in terms of the bias and the Mean Square Error (MSE) of the copula parameters, namely the Gaussian copula. We shall also compare Hoff's method by a non-copula based Kendall's tau computation. To estimate the copula parameters, we combine the Bayesian framework of Hoff (2007) and Pitt et al. (2006). The sampling scheme is separated by first drawing the unknown quantities related to the marginal distributions conditional upon the copula parameters, followed by sampling the copula parameters conditional upon the marginal quantities. It is the first time the effects of different marginal specifications on copula estimates has been studied in a unified Bayesian approach.

We show that by leaving the marginals unspecified, as in Hoff's method, the estimator produces bias which is almost half of the bias from the other estimators. The bias is significant in small samples, but quickly diminishes as the sample size increases. In case of the other estimators the bias persists, and goes down very slowly as the sample increases. The rate is even slower for the parameters capturing dependence involving low count data. The misspecified model (continuous transform) remains the most biased for all the parameters in the Gaussian copula. Hoff's method also has the smallest MSE in all the samples, as compared to the other two methods. The difference in MSE is smaller for the dependence parameter involving discrete data, in small samples. As the sample size increases, the MSE from using Hoff's method decreases at a faster rate, as compared to other specifications. For continuous and high count data, the MSE measure for the dependence parameter through Hoff's method and empirically computed margins, becomes identical. Whereas the misspecified margins produce the largest MSE, regardless of the type of the original margins.

The multivariate modelling scheme through a multivariate copula could very attractive for practitioners, especially when the variables involved are of diverse type. For example using survey data on various individuals, our interest might be to unravel the statistical association between an individuals income, sex, education level, number of children etc. One way would be to obtain some rank-based measures of the bivariate association, but that would require us to deal somehow with the ties observed in the ranks through an ad-hoc method. Secondly, a regression based analysis where a response variable has to be chosen, and as the variables are of different type (binary, ordinal or continuous etc.), the conditional dependence would, lets us say be Gaussian linear regression if income (y_1) is the response variable, or could be a Poisson regression if the response variable is number of children (y_2). In both modelling cases, we don't necessarily get compatible results, as under very specific circumstances we can expect given both conditional distribution $f_1(y_1|y_2, x)$ and $f_2(y_2|y_1, x)$, there to exist a joint probability distribution $p(y_1, y_2|x)$ with f_1 and f_2 as its full conditional distributions. Such a problem persists due to type diversity within the multivariate data, and with lack of knowledge about the joint distribution, generally a response variable is chosen and the analysis proceeds with an appropriate regression, but different results are obtained with different response variables. Hoff's technique tries to address these problems by jointly modelling the variables of interest, and for micro-level data, like that of labour-market, where mixed data types are frequently encountered, an appropriate scale free method is proposed to understand the multivariate association.

In Section 2, we first provide the copula setup and provide details of the various marginal specifications which can be generally employed. In Section 3, we set out the Bayesian sampling scheme for the marginal and the copula parameters. The Data Generating Process (DGP) is explained in Section 4, and in Section 5 we give details of the various marginal specifications used in this paper to estimate the copula. Section 6 will describe the simulation over the DGP, and explain the computation of MSE and bias values. In Section 7 we discuss the results for all the specifications, and conclude in Section 8.

§ 3.2 Gaussian Copula Setup

Our question concerns the effect of various marginal specification $(F_1, \ldots, F_p)$ on the estimates of the copula (chosen) parameters, rather than the effect of marginal specifications across copula families. Therefore we fix a copula function, for the whole analysis. The Gaussian copula is the most frequently employed copula and it offers to model dependence in a linear correlation manner, but without needing to have normal marginals (unlike the Multivariate Normal distribution). The Gaussian copula can be defined as

$$C(u_1, \ldots, u_p) = \Phi_p\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)\},$$

where Φ is the standard normal Cumulative Distribution Function (CDF), and Φ_p is the CDF of a multivariate Normal vector of dimension p. Let us denote a standard normal variable as z_j with zero mean and variance one, which is computed as

$$\Phi(z_j) = F_j^{-1}(u_j), \text{for } j = 1, \ldots, p. \tag{3.2.1}$$

Let $z = (z_1, \ldots, z_p)$, then we can define the Multivariate Normal distribution with zero mean and the covariance matrix equal to the correlation matrix Θ as

$$z \sim N_p(0, \Theta).$$

Song (2000) states that Gaussian copula density equals

$$|\Theta|^{-1/2}\exp(-\frac{1}{2}z'\Theta^{-1}z)\exp(\frac{1}{2}zz'). \tag{3.2.2}$$

(3.2.2) requires the standard normals to be computed through (3.2.1), where F_j is the marginal distribution for the j^{th} component. If F_j is chosen from set of known parametric distributions, then it will have some parameters associated to it. These parameters will need to be estimated along with the Gaussian copula parameters (i.e. correlation matrix). If a non-parametric specification is used, either due to the lack of knowledge about y_j or the limitations a parametric distribution can have, then the corresponding z_j can be obtained through the empirical distribution of the observed data without having to estimate any marginal parameters.

We simplify the problem by not having mixture of marginal specifications in a given multivariate analysis. That is, if F_j is specified to be parametric, then $F_{\setminus j}$ (i.e. all other marginals distributions except F_j) will be parametric as well, and vice versa in the case of non-parametric specifications.

We now present the concept of both parametric and non-parametric marginal distributions, which will later be combined with the Gaussian copula.

3.2.1 Parametric Copula Specification

Let n be the total number of observations given as $y_1, \ldots, y_n$, for $i = 1, \ldots, n$, where each y_i is a $(p \times 1)$ vector. Then the fully parametric Gaussian copula estimation problem is given as

$$\begin{aligned} &z_i \sim N_p(0, \Theta), \\ &y_{ij} = F_{ij}^{-1}\{\Phi(z_{ij})|\beta_j\}, \text{ for all } i \text{ and } j, \end{aligned}$$

where F_{ij} is the CDF for either a continuous or discrete random variable, and β_j is the parameter vector associated with the j^{th} component. For a component j, the marginal distribution F_{ij} is fixed over all the i's, and hence could also be simply stated as F_j. As F_j could be corresponding to either a continuous or discrete random variable, the mapping from y_{ij} to u_{ij} will vary. If the j^{th} component is continuous, F_j^{-1} will be a one-to-one function given β_j, then z_{ij} can be easily be computed. But if the j^{th} component is discrete, F_j^{-1} will be a many-to-one function. Then given β_j, we cannot directly impute the corresponding z_{ij}. We will have to consider them as auxiliary variables, and be sampled along with the copula and the marginal parameters.

Our estimation problem here is similar to Pitt et al. (2006), but we do not account for the presence of covariates in the marginal specification. We could face with the problem of our cross-sectional data having heteroskedasticity which we have to account through our marginal specifications. It is always challenging to model the heteroskedasticity, we could allow for covariates within the marginal distributions which explain it (like square of regressors and cross-products), and then perform i.i.d. tests on the transformed uniform data, like Kolmogorov-Smirnov test. In case of using Hoff's technique, we would have to use a parametric specification to account for the heteroskedasticity first (for example a normal regression) and then perform a test (White test) to ensure no heteroskedasticity in the errors. From there on use the adjusted (fitted) data to proceed with extracting the information through the order statistics for the copula estimates.

3.2.2 Semi-Parametric Copula Specification

If the z_{ij} are computed by assuming a non-parametric marginal distribution, namely an empirical distribution, then along with a parametric copula the estimation problem is on a semi-parametric based specification. In such a setup, there are no marginal parameters which need to be estimated, hence by employing rank based transformations over all the i's for each component j, z_{ij} can be obtained. If all the F_j's correspond to continuous random variables, then an estimator based on the normalized ranks is consistent and asymptotically normal (see Genest et al. (1995)). If however any F_j is a discrete marginal distribution, the ranks are not independent of the marginal distribution F_j, and after dealing with ties in the ranks, such an estimator will be biased. Genest and Nešlehová (2007) show through simulation that a method based on splitting the ties produces the smallest bias in the estimation of Θ. For discrete data, the functional form of the marginal distribution is not independent of the dependence measure.

Hoff (2007) presents a semi-parametric copula estimation technique, which unlike the method explained above, treats all the z_{ij} as auxiliary variables. No assumption is made regarding F_j, and it is treated as completely unknown. He proposes a likelihood, which is independent of the nuisance marginal parameters, and estimates the multivariate Gaussian copula parameters for diverse data types. Hoff's method is applicable to discrete, continuous or mixture of both data types. For continuous margins, it produces similar results as an estimator based on the rank likelihood.

3.2.2.1 Empirical Distribution $\widetilde{F}_j$

If empirical distributions are used for all the marginals in a multivariate Gaussian copula, then there are no parameters associated to any components. Then the modelling problem becomes

$$\begin{aligned}
&z_i \sim N_p(0, \Theta),\\
&y_{ij} = \widetilde{F}_{ij}^{-1}\{\Phi(z_{ij})\},\\
&\widetilde{F}_j(y_{mj}) = \tfrac{1}{n+1}\textstyle\sum_{i=1}^{n} \mathbb{1}(y_{ij} \le y_{mj}), \text{ for all } i \text{ and } j.
\end{aligned}$$

$\widetilde{F}_j$ denotes the empirical distribution, used instead of a parametric F_j for all j, the division of $n+1$ is to avoid boundary cases. We only need to estimate the correlation matrix Θ, and in case any of the random variable is discrete, then the corresponding $\widetilde{F}_j^{-1}$ is a many-to-one function. Analogues to Genest et al. (1995), where they spilt the ties in the observed data before proceeding with copula estimation, we uniformly sample them from the interval given through the empirical step-size. Hence they are also considered unknown over a fixed uniform interval.

3.2.2.2 Unknown F_j

Here, unlike employing an empirical CDF for all the marginal distributions, we treat all F_j's as completely unknown, and therefore do not know z. The only information we have regarding F_j is that for all the components they are non-decreasing functions. We can also determine the corresponding rank for each observed y_{ij}. If the rank of y_{ij} is k, then the order statistic of y_{ij} is $y_j^{(k)}$, such that $y_{ij} = y_j^{(k)}$. From this information we can infer that the unobserved z_{ij} corresponding to y_{ij}, will have the same rank k, and can be written formally as

$$y_j^{(k-1)} < (y_{ij} = y_j^{(k)}) < y_j^{(k+1)}, \text{implies}, \tag{3.2.3}$$

$$z_j^{(k-1)} < (z_{ij} = z_j^{(k)}) < z_j^{(k+1)}. \tag{3.2.4}$$

From (3.2.4), we know for certain that z_{ij} has to lie in the interval dictated by the order statistics of the observed data. Based on this information, we set out the same Gaussian copula specification as

$$z_i \sim N_p(0, \Theta),$$

$$y_{ij} = m, \text{ if } \max\left\{z_{rj};\ F : m-1 \mapsto z_{rj}\right\} < z_{ij} < \min\left\{z_{rj}; F : m+1 \mapsto z_{rj}\right\}, \text{ for all } i \text{ and } j,$$

$$\text{where } m \in M \ \text{(discrete outcomes)}.$$

In the case of continuous margins and large samples, the interval where z_{ij} lies in becomes smaller, and hence the uncertainty regarding the true value of z_{ij} is reduced. In that case the methodology is similar to assuming an empirical distribution for the respective margin $\widetilde{F}_j$. This method of specifying the marginals uses the only certain information, which is given through the order statistic due to the non-decreasing monotonic property of the unknown F_j.

In the next section we describe the bayesian sampling scheme to estimate Θ, for the marginal specifications described above.

§ 3.3 Bayesian Estimation

Our aim is to estimate the Gaussian copula parameters, namely the correlation matrix Θ. If all the margins are parametrically specified, then each component j will have a parameter vector β_j associated to it. Let β denote the vector containing all the marginal parameters from all the components, $\beta = (\beta_1, \dots, \beta_p)$. In a parametric setting, if any of the random variables is discrete, we also need to sample the unknown standard normals z_{ij}, for that component. In case of non-parametric marginal distributions, there are no marginal parameters to be estimated, but only Θ. If an empirical distribution is specified for a continuous random variable, then the z_j corresponding to it are simply computed through the inverse empirical CDF, but in case of discrete outcomes, it will be sampled through the empirical step-size to deal with ties. Through Hoff's method, regardless of the type of margin involved, the copula parameters and z have to be computed conditional upon each other.

Given these requirements and settings, we can partition the bayesian sampling scheme into two parts. First $\beta = (\beta_1, \ldots, \beta_p)$ (for parametric margins) and $z = (z_1, \ldots, z_p)$ (if needed) are sampled conditional upon Θ. Secondly, we sample Θ conditional upon β and z.

3.3.1 First Stage $p(\beta, z|\Theta)$

In the first stage, we sample the quantities related through the marginals which are needed to estimate the copula parameters. In case of non-parametric marginals, the conditioning probability is $p(z|\Theta)$, without β.

3.3.1.1 Parametric Margins

In case the marginals are all parametrically specified, then we sample in this order

1. Sample from $p(\beta_j|y_{.,j}, z_{.,\backslash j}, \Theta)$, where $y_{.,j}$ denotes all the observations n for the given component j, and $z_{.,\backslash j}$ denotes all the observations from all the other components except j.

2. If j^{th} marginal distribution F_j is continuous, then compute $z_{ij} = \Phi^{-1}\{F_{ij}(y_{ij}|\beta_j)\}$. If F_j is a discrete distribution, we sample z_{ij} from $p(z_{ij}|\beta_j, y_{ij}, z_{i,\backslash j}; \Theta)$, for all i.

The above two steps are repeated for each j, in turn. Pitt et al. (2006) provide details about the conditional density of β_j, from which it is not always possible to sample from directly. They propose a Metropolis-Hasting algorithm, where the proposal density is approximated to a multivariate t-distribution, with mean equal to the mode $\widehat{\beta}_j$ of log $p(\beta_j|y_{.,j}, z_{.,\backslash j}, \Theta)$. The mode can be found through numerical methods like quasi-Newton Raphson method. The variance of the t-distribution is equated to the negative inverse of the second derivative of the log conditional density, computed at the mode. The degrees-of-freedom is arbitrarily chosen such that the proposal density can dominate the true density in the tails. Such a method is similar to a Laplace-type proposal (see Chib and Greenberg (1998) and, Chib and Winkelmann (2001)). A new proposed value β_j^* is then evaluated in a Metropolis-Hasting step.

In case a component j has a discrete marginal distribution, we first sample β_j, and then conditional upon it z_{ij} are sampled from a truncated univariate Normal distribution, where the mean and the variance are equated through the correlation among the other components $z_{i,\setminus j}$. We refer the interested reader for full details of the briefly explained above algorithm to Pitt et al. (2006), page 542-544.

3.3.1.2 Unspecified Marginals

If a semi-parametric copula approach is adopted, where no assumption regarding F_j is made, then there is no β_j to be sampled, but only z needs to be sampled in this stage. In case an empirical distribution is assumed for a discrete random variable, then z_{ij} corresponding to the observed y_{ij} is sampled uniformly through the interval

$$\Phi(z_{ij}) \sim [\widetilde{F}_j(y_{ij}-1), \widetilde{F}_j(y_{ij})], \text{ for all } i \text{ and } j,$$

where $u_{ij} = \Phi(z_{ij})$. The full set of z is obtained, and then we proceed on to the second stage. This implies we are splitting the ties in random dictated by the empirical-step size, as suggested by Genest and Nešlehová (2007).

To employ the approach set out by Hoff (2007), we need z_{ij} to be sampled from

$$z_{ij} \sim p(z_{ij}|\Theta, z_{i,\setminus j}, y_j^{(k)}), \text{ for all } i \text{ and } j,$$

where the conditional density of z_{ij} is conditioned on the correlation matrix Θ and all the other standard normals from each j. The conditioning of $y_j^{(k)}$ implies z_{ij} has to lie in the interval $[z_j^{(k-1)}, z_j^{(k+1)}]$, that is it has to obey the order statistics. Hoff (2007) specifies a full conditional of z_{ij}, which is a truncated univariate Normal distribution with mean and variance accounting for correlation between other other components $z_{i,\setminus j}$. Each i for a fixed j is sampled in turn. The major difference in sampling the z here is that the truncation is dictated by the order statistics, whereas in the discrete parametric case, the truncation is given by the CDF, evaluated at y_{ij} and $y_{ij}-1$ (see Pitt et al. (2006)). This scheme is invariant to either discrete or continuous margins. The full details of the sampling of z here, are provided in Hoff (2007), page 273.

3.3.2 Second Stage

In this stage, we no longer care about the assumptions specified on the marginal distributions. All we require are z from the previous stage, to sample Θ. Hence, either parametrically defined marginal distribution or non-parametrically, this scheme for Θ is invariant. We can write the posterior of Θ as

$$p(\Theta|z) \propto p(\Theta) \times p(z|\Theta).$$

Similar to Hoff (2007), we assume a semi-conjugate prior for the Gaussian copula. The prior $p(\Theta)$ is defined through V, which is specified to have a prior given as an inverse-Wishart distribution $(\nu_0, \nu_0 V_0)$, parametrized such that $E[V^{-1}] = V_0^{-1}$, where ν_0 is the degrees-of-freedom and $\nu_0 V_0$ the scale matrix. Θ is equated as

$$\Theta_{[i,j]} = \frac{V_{[i,j]}}{\sqrt{V_{[i,i]} V_{[j,j]}}}.$$

The posterior of V can then be shown to be proportional to

$$V|z \sim \text{inverse-Wishart}(\nu_0 + n, \nu_0 V_0 + z'z),$$

from which a sample of V can be obtained, and then Θ computed from the above transformation.

We follow Hoff (2007) rather than Pitt et al. (2006) to specify the prior of Θ, because our focus is not on covariance selection methods, but to check the effects of the marginal specifications on copula estimation.

§ 3.4 Data Generating Process

In this section we explain how to simulate data from a multivariate Gaussian copula and provide details about the Data Generating Process (DGP). The simulated data will be used to test various marginal specifications and their effect on a Gaussian copula estimation. For some correlation matrix Θ and marginal parameters β, a set of generated y can be sampled as follow

1. Sample z from $N_p(z|0, \Theta)$.

2. Obtain $u = \Phi(z)$.

3. Compute $y_{ij} = F_j^{-1}(u_{ij}|\beta_j)$, for all i and j.

Where $u = (u_1, \ldots, u_p)$, and each u_i is $(n \times 1)$ vector. Step 3 above implies, that we need to be able to compute the inverse CDF of all the chosen parametric marginal distributions. Let us then set out the DGP, which will be used throughout the rest of the paper. We choose $p = 3$ and alter n such that it ranges from small sample ($n = 10$) to large sample ($n = 500$).

$$z \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} 1 & 0.8 & 0.4 \\ 0.8 & 1 & 0.6 \\ 0.4 & 0.6 & 1 \end{pmatrix}\right],$$

$$\begin{aligned}
&u = \Phi(z), \\
&y_{.,1} = F_1^{-1}(u_{.,1}|1.5) \Rightarrow F_1(y_{.,1}|1.5) = \text{Exponential}(y_{.,1}|\lambda_1), \\
&y_{.,2} = F_2^{-1}(u_{.,2}|6) \Rightarrow F_2(y_{.,1}|6) = \text{Poisson}(y_{.,2}|\lambda_2), \\
&y_{.,3} = F_3^{-1}(u_{.,3}|0.6) \Rightarrow F_2(y_{.,1}|0.6) = \text{Bernoulli}(y_{.,2}|\lambda_3).
\end{aligned}$$

So the true DGP is a mixture of continuous and discrete marginals. This DGP will stay fixed throughout the simulation and we will assume various marginals specifications to estimate the correlation matrix Θ.

§ 3.5 Marginal Specifications

Now we state the various marginal specifications we will employ, in order to estimate the Gaussian copula parameters. For ease of reference, we can refer to them as Marginal Specifications (MS), so various specifications will be defined as MS1, MS2 etc. Their detail is as follow:

- **MS1** All three marginals (F_1, F_2 and F_3) are assumed to be completely unknown. Using the order statistics of the observed data, first z, and then the correlation matrix is sampled. This is as described previously and is the method proposed by Hoff (2007).

- **MS2** Use empirical CDF $\widetilde{F}_j$ for all three margins, and compute $z_{ij} = \Phi^{-1}(u_{ij})$, where u_{ij} is uniformly sampled from the interval $[\widetilde{F}_j(y_{ij} - 1), \widetilde{F}_j(y_{ij})]$,

- **MS3** Make the two discrete margins continuos by adding a random $[0, 1]$ error, then let $z_{ij} \sim \ln \mathcal{N}(y_{ij}|\mu_j, \sigma_j)$, for all i and j. Hence all margins are log normally distributed.

So we only specify three different marginal specifications. The first two correspond to semi-parametric copula estimation, and the last to a fully parametric copula estimation. We decided to consider misspecified margins, as in very small sample it is interesting to see how well they perform in estimating the copula parameters. MS3 takes the discrete marginals and adds a uniformly $[0, 1]$ random noise to the observed values, to make them continuous. This is an approach stated in Trivedi and Zimmer (2006), to avoid computational problems generally encountered in likelihood estimation. This transformation along with assuming log normal distribution induces a misspecification. The first margin (originally exponential in the DGP) is also misspecified by assuming a log normal.

Next, we look at the simulation over the DGP in more detail.

§ 3.6 Simulation

3.6.1 Setup

The sampling scheme previously described is a kind of a Gibbs type sampler over the two stages defined. To obtain the posterior density of Θ, we perform 6000 iterations from which every 5th iteration is saved. After thinning, the autocorrelation within the chains computed through all the specifications is very low. Further by dropping the first 200 iterations for burn-in, it becomes insignificant after few lags. So the final sample size to conduct posterior inference is 1000. Our quantity of interest is $\mathrm{E}(\Theta|y)$ through all the marginal specifications. To analyse the properties of the various marginal specifications and their effect on the estimation of Θ, we have to obtain a distribution for the posterior mean itself, hence we employ Monte Carlo over the DGP. The size of the Monte Carlo simulation is 250, which is sufficient as convergence for the quantities computed is quick. At each Monte Carlo iteration we obtain a new sample of y through the same DGP, which can be denoted as $\{y\}_s$, where $s = 1, \ldots, S$. We can define the general simulation structure as,

for $s = 1, \ldots, S$,
sample $\{y\}_s$ from the DGP,
obtain $\mathrm{E}\Big[\Theta|\{y\}_s\Big]$, for all MS1, MS2 & MS3.

The above scheme is repeated for various sample sizes, $n = 10, 25, 50, 100, 250, 500$.

3.6.2 Bias & Variance

After obtaining the distribution of the posterior mean of Θ for all the three marginal specifications, we compare them in terms of their bias and variance towards the true correlation matrix Θ_T, defined in section 4. We compute two quantities of interest for all the marginal specifications. First, we compute the Bias through the difference of $\mathrm{E}\Big[\Theta|\{y\}_s\Big]$ from Θ_T. Secondly, we compute the Mean Square Error (MSE), which combines the variance and bias of the estimator. These are given as

$$\text{Bias} = \frac{1}{S}\sum_{s=1}^{S}\mathrm{E}\Big[\Theta|\{y\}_s\Big] - \Theta_T,$$

$$\text{Mean Squared Error (MSE)} = \frac{1}{S}\sum_{s=1}^{S}\Big[\mathrm{E}[\Theta|\{y\}_s] - \Theta_T\Big]^2.$$

We will compare the bias across all three methods (marginal specifications), and as our interest is in determining the performance of MS1 compared to the other specifications, we will compute the MSE ratio of MS1 with respect to MS2 and MS3

$$\omega_{12} = \frac{\text{MSE}_{\text{M1}}}{\text{MSE}_{\text{M2}}}, \qquad \omega_{13} = \frac{\text{MSE}_{\text{M1}}}{\text{MSE}_{\text{M3}}}.$$

These quantities will be computed for the whole correlation matrix Θ.

§ 3.7 Results

3.7.1 Bias

We present the results from computing the bias through using all the marginal specifications in Table 3.1. $\Theta_{[.,.]}$ represents a parameter capturing the dependence between two particular random variables. The bias from Hoff's method B_{MS1} is lower in all sample sizes and for all parameters, as compared to B_{MS2} (bias from empirically computed margins) and B_{MS1} (bias from misspecified margins). For sample size of $n = 10$, we see the bias B_{MS1} for all the parameters is almost half that off B_{MS2} and B_{MS2}. The lower bias is particularly noticeable for $\Theta_{[1,3]}$ (correlation between exponential and a binary variable) of -0.1137, which is one-third of the bias from the other estimators. There are only two ranks for the binary variable and the z are sampled through these ranks in MS1, hence considering all the uncertainty. The misspecified model produces similar bias as the empirically specified model for $n = 10$, which asserts the point that even a misspecified model can be used instead of MS2 for very small sample sizes. As n increases to 25, the bias from MS1 drops by half for all the parameters. The bias from $\Theta_{[1,2]}$ in the case of MS2 also drops by half, but for $\Theta_{[1,3]}$ and $\Theta_{[2,3]}$ (correlation between poisson and binary variables) the bias reduces slightly, this is also true for the bias from MS3. Through increasing n, we see the bias in MS1 further reduces, and the rate of reduction is faster as compared to MS2 for correlation parameters involving the binary variable ($\Theta_{[1,3]}$ and $\Theta_{[2,3]}$). The bias in $\Theta_{[1,2]}$, which is the parameter denoting the correlation between a continuous and high count data is almost equal for MS1 and MS2 in large samples, which emphasis the appropriateness of empirically computed margins for continuous data and high count data. The bias from MS3 decreases at a slower rate as compared to MS2, as n increases, which shows that using misspecified margins (transforming discrete data) is not appropriate, and will produce wrong results. Overall, we see MS1 produces smaller bias as compared to the other estimators, and it especially performs well for measuring dependence among discrete data. The information contained within the order statistics and having a likelihood conditional upon this ensures an unbiased estimate for the copula parameters of interest.

Table 3.1: *Bias for all Marginal Specifications*

	n=10	n=25	n=50	n=100	n=250	n=500
B_{MS1}						
$\Theta_{[1,2]}$	-0.2074	-0.0794	-0.0509	-0.0312	-0.0183	-0.0108
$\Theta_{[1,3]}$	-0.1137	-0.0334	-0.0319	-0.0120	-0.0127	-0.0053
$\Theta_{[2,3]}$	-0.2177	-0.0840	-0.0639	-0.0349	-0.0222	-0.0119
B_{MS2}						
$\Theta_{[1,2]}$	-0.4153	-0.2041	-0.1184	-0.0622	-0.0352	-0.0245
$\Theta_{[1,3]}$	-0.2891	-0.2123	-0.2021	-0.1642	-0.1602	-0.1539
$\Theta_{[2,3]}$	-0.3797	-0.2658	-0.2278	-0.1910	-0.1725	-0.1639
B_{MS3}						
$\Theta_{[1,2]}$	-0.3985	-0.2389	-0.1907	-0.1525	-0.1358	-0.1262
$\Theta_{[1,3]}$	-0.2731	-0.2263	-0.2136	-0.2086	-0.1985	-0.1962
$\Theta_{[2,3]}$	-0.3995	-0.3084	-0.2990	-0.2919	-0.2790	-0.2775

3.7.2 MSE

The MSE ratio results for MS1 against the other two specifications are reported in Table 3.3 for various n. For $n = 10$, we see the ratio ω_{12} (for Hoff's method against the empirical specification) for all the parameters is less than one, indicating that Hoff's method MS1 produces smaller variance as compared to MS2. This is also true in comparison to MS3 in ω_{13}, and the ratio is quite similar to those of ω_{12}, which is indicating again that for small n we can somehow mispecify and still get reasonable results. We see the MSE ratio for the continuous and poisson variable dependence parameter is lower compared to other bivariate random variables estimates, this indicates even though the sample is small, still Hoff's method's estimates are close to the true ones. When one of the variables involved is a binary variable, the uncertainty is large due to only two ranks available, and therefore the ratios ω_{13} and ω_{23} are large in small samples, implying the gain in efficiency from Hoff's method is not that substantial. As n increases, we see the MSE ratio for MS1 against the other two becomes much lower, which implies even though we are dealing with highly discrete data, but with large n we can still on average estimate the parameters more efficiently through Hoff's method. More information is present in large n about the true dependence, which MS1 captures. The ratio ω_{12} for $\Theta_{[1,2]}$ increases with n, indicating that computing margins through an empirical distribution for continuous and high count data in large n will become quite similar to Hoff's method in terms of efficiency. This is leaning towards the usage of an empirical distribution for almost continuous like variables. But when one of the variable is binary type, the variance ratio ω_{12}, starts dropping as n increases and gets close to zero ($n = 500$). Which implies MS1 is more efficient compared to other specifications. For using misspecified margins MS3, the MSE ratio drops at a much faster rate as compared to MS2, and for $n = 500$, we see the ratios are all close to zero (for all margins) which implies the bias created through the addition of the noise keeps the estimates away from the true parameters. The sampling of z in the empirical case does consider some uncertainty, but still it has fixed step-size interval. Whereas in Hoff's method such interval for a specific rank can expand and retract and hence can consider much more uncertainty, therefore leading to more efficient estimates. An interesting point to note is also how for large n, MS2 and MS3 have similar efficiency in case of low count data, which does somehow makes a case for continuous transformation as suggested in the literature.

Table 3.2: *MSE ratio*

	n=10	n=25	n=50	n=100	n=250	n=500
ω_{12}						
$\Theta_{[1,2]}$	0.3633	0.3203	0.3885	0.4681	0.5031	0.4275
$\Theta_{[1,3]}$	0.7640	0.6181	0.4513	0.3107	0.1854	0.0708
$\Theta_{[2,3]}$	0.6550	0.4334	0.3295	0.2098	0.1225	0.0531
ω_{13}						
$\Theta_{[1,2]}$	0.3838	0.2374	0.1580	0.0950	0.0461	0.0233
$\Theta_{[1,3]}$	0.8354	0.5694	0.4184	0.2036	0.1249	0.0442
$\Theta_{[2,3]}$	0.5725	0.3179	0.2033	0.0958	0.0483	0.0188

3.7.3 Non-Copula Alternative

As we are interested in computing the correlation among random variables of different types, we could adopt a simpler approach to do so. The closest method in terms of retrieving information through the order statistics would be to compute the Kendall's tau (τ) rank correlation. It is computed as

$$\tau = \frac{\text{(no. of concordant pairs)}-\text{(no. of discordant pairs)}}{\frac{1}{2}n\,(n-1)}.$$

It accounts for the ties observed in the ranks, which is vital for discrete outcomes. Of course there is no need to specify any marginal density or the need of any copula family. We simply take the simulated data and compute τ for all three bivariate correlation pairs. Comparison will be made with the results from Hoff's method again in terms of the bias ($\mathrm{B_K}$) and MSE ratio ($\omega_{1\mathrm{K}}$).

Table 3.3: *Kendall's rank correlation (bias and MSE)*

	n=10	n=25	n=50	n=100	n=250	n=500
$\mathrm{B_K}$						
$\Theta_{[1,2]}$	0.3595	0.8228	0.5260	0.4668	0.4950	0.4494
$\Theta_{[1,3]}$	-0.2783	0.4297	0.1292	0.1798	0.1350	0.1060
$\Theta_{[2,3]}$	-0.4557	0.4502	0.2210	0.3101	0.1689	0.1800
$\omega_{1\mathrm{K}}$						
$\Theta_{[1,2]}$	1.0085	0.3936	0.1521	0.0671	0.0225	0.0087
$\Theta_{[1,3]}$	0.9199	0.9130	0.6380	0.3881	0.2254	0.1814
$\Theta_{[2,3]}$	1.1954	0.6230	0.3902	0.1983	0.0992	0.0399

When the number of observations are really low ($n = 10$), we see that bias is still higher for correlation through Kendall's tau as compared to Hoff's method, but the MSE is almost one. Unlike MS2 and MS3, kendalls' tau has equal variance for small n. As n increase, we see the bias for all n in case of kendall's tau persists and is similar to MS2 and MS3. But compared to Hoff's method the bias is much larger and has a slower rate of reducing. For the MSE ratio, we also see that the ratio comes closer to zero, which is indicating again that Hoff's method has smaller variance as more data becomes available. In terms of the MSE, the ratio is similar to that of MS3, which is the misspecified marginal specification.

Hence we overall see that computing correlation through Kendall's tau produces higher bias and larger variance as compared to Hoff's method. Against MS2 and MS3, the bias is similar but in terms of efficiency it is similar to the misspecified model MS3. Therefore to avoid bias we should Hoff's method, but to adopt a simpler method one can compute correlation through Kendall's tau, rather than MS2 and MS3 at the cost of loss in efficiency.

3.7.4 MS1 Kernel Density

We present some kernel density plots for the first marginal specification, assuming F to be completely unknown. Figure 3.1-3.5 are density plots for the posterior mean over the DGP. We can clearly see the dispersion around the mean (dotted line) decreases as n increases, and also the distance to the true parameter value (full line) reduces. We can see that the dispersion of the posterior mean $\mathrm{E}[\Theta_{[1,3]}|y_s]$ and $\mathrm{E}[\Theta_{[2,3]}|y_s]$ is relatively higher as compared to that of $\mathrm{E}[\Theta_{[1,2]}|y_s]$, through all sample sizes, which is as described before due to the uncertainty created by binary variable involved. Density plot $\mathrm{E}[\Theta_{[1,3]}|y_s]$ and $\mathrm{E}[\Theta_{[2,3]}|y_s]$ for small n show wide dispersion, almost stretching through the whole correlation parameter space $[-1, 1]$. As n increases, the dispersion for all the parameters gets close to the mean of the posterior means (dotted line).

Figure 3.1: *Posterior Density of* $E(\Theta|y)$, $n = 10$.

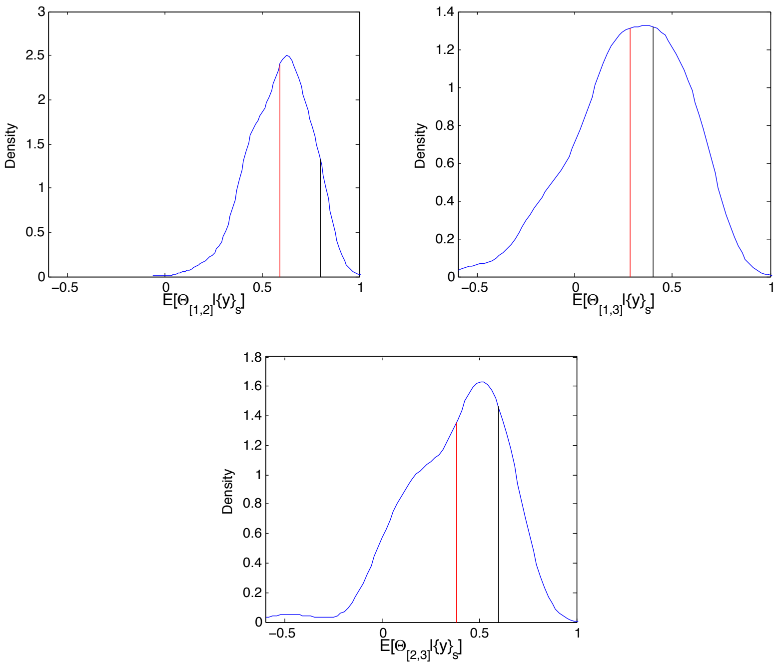


Figure 3.2: *Posterior Density of* $E(\Theta|y)$, $n = 25$.

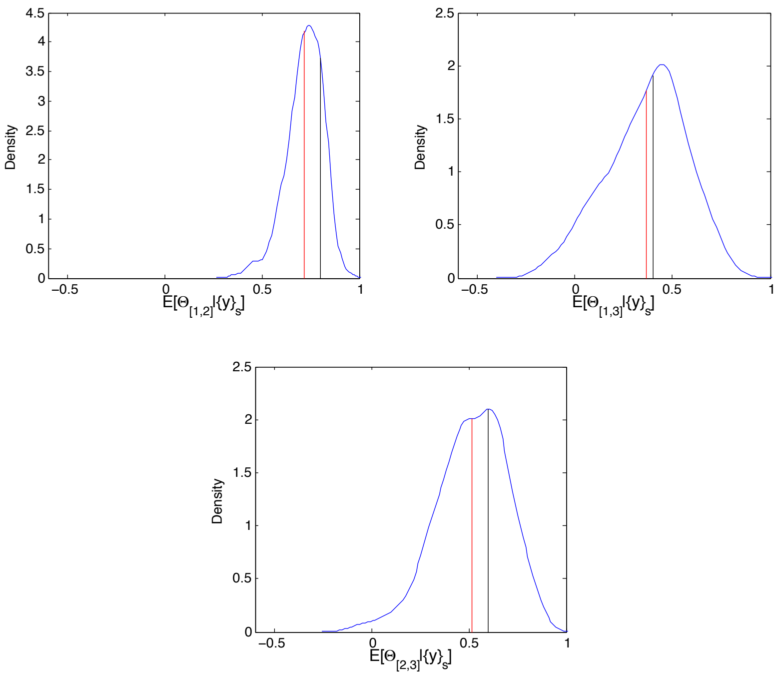


Figure 3.3: *Posterior Density of* $E(\Theta|y)$, $n = 50$.

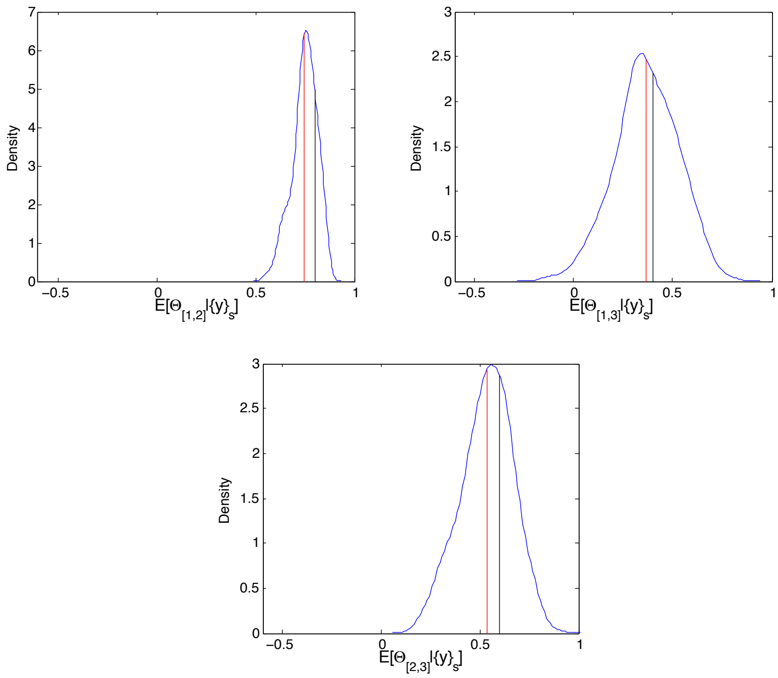


Figure 3.4: *Posterior Density of* $E(\Theta|y)$, $n = 100$.

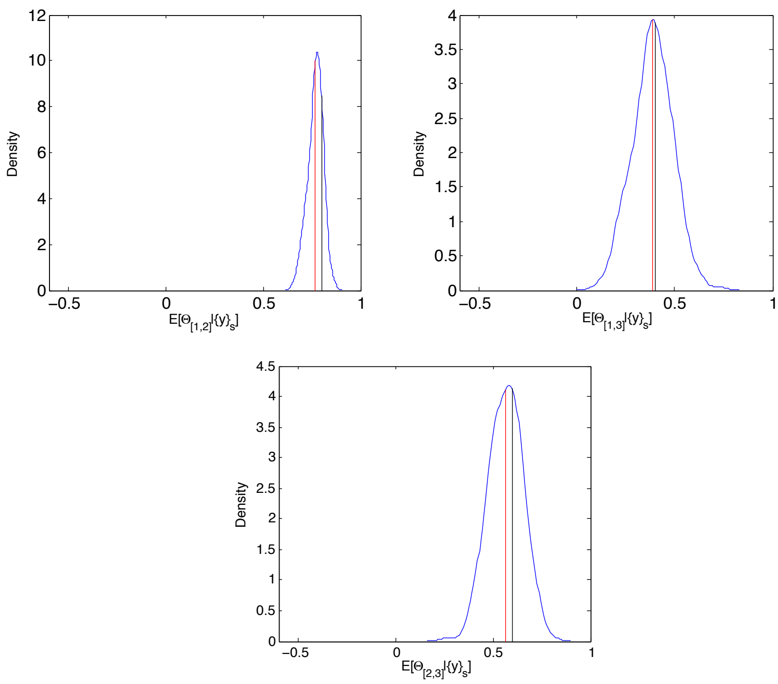


Figure 3.5: *Posterior Density of* $E(\Theta|y)$, $n = 250$.

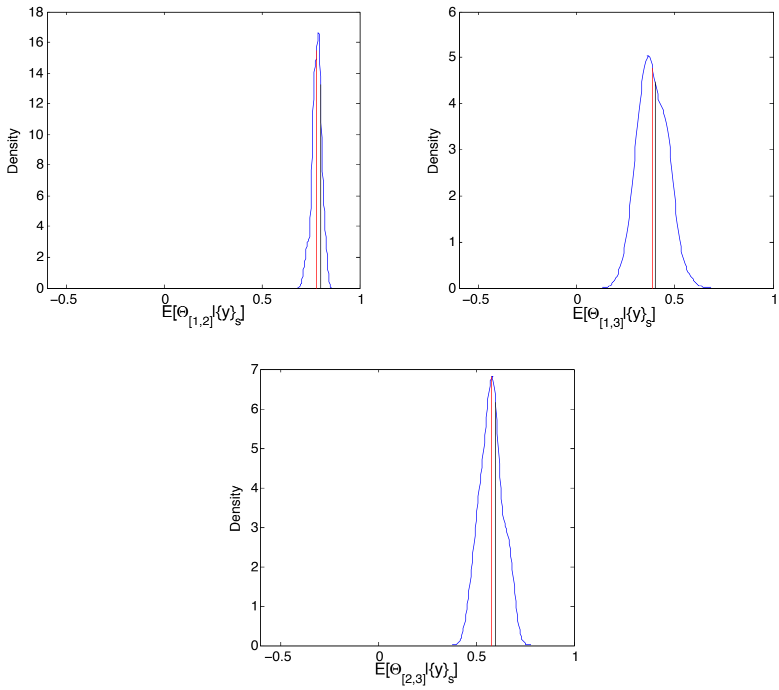
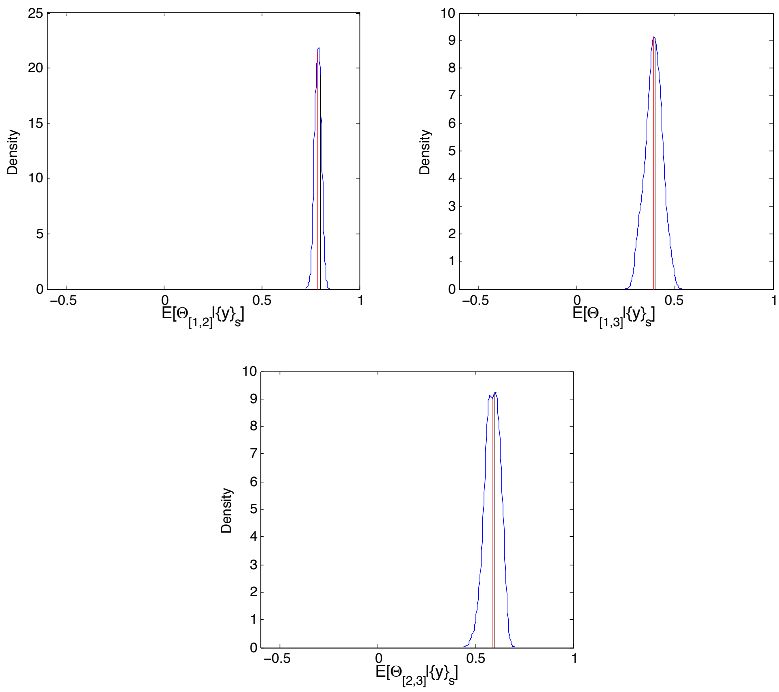


Figure 3.6: *Posterior Density of* $E(\Theta|y)$, $n = 500$.



§ 3.8 Conclusion

Copula based method is a flexible method to conduct multivariate analysis for margins of different types and dependence patterns which are not best described by elliptical distributions. When marginals are specified parametrically along with a parametric copula, then the estimation problem is fully parametric. Alternatively, a non-parametric distribution can be used for the marginals which lead to a semi-parametric copula estimation problem. For random variables of continuous type, a semi-parametric copula estimation is shown to be efficient and asymptotically normal (see Genest et al. (1995)), but for multivariate analysis of discrete or mixture of continuous-discrete data, empirically computed marginals are not appropriate. Even in the parametric case, the exact knowledge of the marginal distribution is not always available and transformation to a continuous distribution is also not fruitful. Certain type of marginal distributions also do not belong to standard parametric families. Hoff (2007) proposes a method, where the marginal parameters are not required to be estimated and by simply obtaining the information contained in the order statistics, we can estimate the copula parameters. Such a method is useful, as it allows combining marginals of all different types and does not suffer from any possible misspecification.

In this paper, we evaluated the effect on a Gaussian copula estimation due to various marginal distribution specifications. In particular, we study the approach of Hoff (2007) where the marginals are left completely unspecified. Copula estimation is performed in a full Bayesian framework of Pitt et al. (2006) and Hoff (2007). Apart from Hoff's method, one approach is where the margins were empirically computed, and another where we completely misspecified the margins before estimating the Gaussian copula.

The results showed that Hoff's method outperforms the other two specifications in all sample sizes. It produces the smallest bias, and for correlation estimates between different data types the bias quickly approaches zero in large samples. Using empirically computed margins, produces similar bias for continuous data, but for discrete data the bias reduces at a slower rate as compared to Hoff's method. The misspecified margins produce similar bias as empirically computed margins in very small samples, but as the sample increases it becomes more inappropriate to use such a method. In terms of MSE, again Hoff's method outperforms the other two specifications. In small sample, the MSE is similar for correlation estimates of discrete data, but the ratio approaches zero as the sample size increases. For the case of continuous and high count data, the MSE ratio between Hoff's method against empirically computed margins increases as the sample size increases. In case one of the random variable is a binary type, the MSE through misspecified margins is similar to the empirically computed ones. But through Hoff's method the MSE decreases in relative terms compared to other two. We also compared Hoff's method to a non-copula based approach, as that presents an ease in computing the correlation.

We could also use multivariate copula to make predictions over a variable given other variables are given. In which case we would require the availability of the full conditional distribution, for instance in our case of a Multivariate Gaussian copula, given knowledge of the marginal distributions and the covariance matrix, we can can obtain predictions of the unknown variable. However, through Hoff's method it is not straightforward, as it would require some arbitrary assumption on the marginals. This still being possible, predictions over random variables given other variables is not of the attractive usages of copula functions.

Overall, even though for continuous data and large samples both empirical and Hoff's method are equivalent, but for discrete data, Hoff's method performs better than empirical or misspecified margins in all sample sizes. For a multivariate analysis of diverse data types, estimation based on Maximum Likelihood can run into many problems, and Bayesian techniques offer alternative to address the problems associated with discrete data types. Hoff's method not only addresses the issue of discreteness, but avoids any possible misspecification of the marginals. The method can also be adopted to other copula families, through an appropriate Markov Chain Monte Carlo technique.

- Chapter 4 -

Bayesian Inference for a Semi-Parametric Copula-based Markov Chain

§ 4.1 Introduction

Limitations and rigidness of time series models is well documented. Their construction has always depended upon the exact nature of the data, and do not easily accommodate other types of data (non-normal random variables). By separating the characteristics specific to the data from the time-varying properties, we are able to specify a general method through copula to model any strictly stationary time series. Such a copula-based Markov chain is applicable to discretely varying data (i.e. binary, count or ordered data) and also to the case of continuous random variable. The advantage of such a technique is, first, we are able to separate out the marginal behaviour of a time series from the dependence structure (like asymmetric dependence and tail dependence). Secondly, using tail-dependent copulas we can specify a time series process which acts like long memory, usually encountered in financial and economic applications.

Most of the work on copula-based Markov chain deals with the theoretical aspects related to it, like probability and weak dependence properties. Darswo and Olsen (1992) specify the necessary and sufficient conditions for constructing stationary first-order Markov models based on a copula. They show that the Chapman-Kolmogorov equations are satisfied for such models. Ibragimov (2009) extend the conditions presented by Darswo and Olsen (1992) for higher-order Markov models. Chen and Fan (2002), Chen et al. (2009) and Beare (2010a,b) present the persistence properties of stationary copula-based Markov chains. Chen and Fan (2002) state conditions under which copula-based Markov models are β-mixing with either exponential or polynomial decay rates, and the conditions are independent of the marginal distribution specification, but only dependent upon the copula specification, they also show EFGM and Gaussian copula are indeed geometric β-mixing. Beare (2010a) provides strong sufficient conditions for geometric β-mixing, which rules out copula families exhibiting asymmetric and tail dependence, for whom β-mixing with exponential decay rates is established. Lentzas and Ibragimov (2008) show a Clayton copula-based model behaves like a long memory time series with high persistence, but Chen et al. (2009) show in terms of the mixing properties, such a model is weakly dependent and short memory and models generated through Clayton, Gumbel and t-copula are indeed geometric β-mixing. Beare (2010b) shows for Archimedean copulas that the regular variation of the generator at zero and one implies geometric ergodicity. Joe (1997) shows in a fully-parametric copula setting that various Maximum likelihood (ML) based estimators to be consistent and asymptotically normal under some regularity conditions. Chen and Fan (2002) propose a semi-parametric copula estimation (empirically computed margins) using ML, and also prove consistency and asymptotic normality. Chen et al. (2009) state an efficient sieve ML estimation procedure for copula-based Markov chains.

Previously the copula-based Markov chain literature has been restricted to case of continuous marginal distributions. For the first time, we are proposing a method to model discretely-varied time series through copulas, where we address the difficulties generally faced in discrete data modelling. Our method is also directly applicable to data of continuous type, as we make no assumption regarding the data (similar to Hoff (2007)). The paper opens the literature in applying Bayesian techniques to estimate a copula-based Markov chain. Recently Bayesian techniques have provided solutions to problems posed by non-continuous margins in copula modelling (see Pitt et al. (2006) and Smith and Khaled (2012). Most time series models lack long memory features generally observed in economic time series, like period of high unemployment are likely to be followed by further periods of high unemployment. Our methodology not only allows to capture such persistence, but it is also flexible to different type of dependence patterns.

The modelling is similar to an AR type process, where we specify the current value to be some function of its own lags. But unlike previous models, where the assumption on the marginal distribution dictates the conditional distribution (normality etc.), we model these distributions separately and they are not bounded to each other. Our method is general across various data types, as we make no assumption regarding the marginal distribution and treat it completely as unknown. The estimation of the copula parameters (conditional dependency) is based only on the order statistics of the observed time series which is similar to Hoff (2007), but he deals with cross-sectional data, and specifies a sampling scheme suitable only to a Gaussian copula. To keep the intuition clear, we model a first-order Markov chain, which can be adopted for high order processes. The marginal distribution is completely left unspecified, and we treat the uniform variables (generally obtained through the marginal distribution) as latent variables, which along with the copula parameters are estimated in a Bayesian framework.

We use a real data application, which is based on the count of weekly firearm homicides observed in Cape Town, South Africa. Crime in general is quite persistent, and our period of analysis consists of a time when there was urbanisation in and around Cape Town. Hence we could model such persistent through copulas. We successfully capture temporal dependence first through a Gumbel copula, and then to show our method is invariant to different copula families (also applying a Gaussian copula separately). In terms of the Bayesian methods applied, standard diagnostics are used to confirm the Markov Chain Monte Carlo (MCMC) performs well.

The paper starts with setting out a copula framework, where we cover vital aspects related to the literature on copula-based Markov process like necessary conditions, mixing properties and estimation methods. In Section 2, we set out the necessary framework and the modelling problem. We then set out a two-stage Bayesian sampling scheme for the latent copula arguments and the copula parameters in Section 3. Section 4 presents a real-data application of the technique specified and then concluding with some ongoing extensions to the method.

§ 4.2 Copula-based Time Series (Review)

This section gives a detailed review on copula-based Markov chain literature, which has only recently gained popularity. Primarily, we cover the conditions needed for a specifying a copula-based Markov process, covering their mixing properties and providing a summary for various estimation methods.

4.2.1 Copula-based Markov chain

Most of the copula literature deals with modelling the dependence between two random variables (X and Y above), namely the contemperouns dependence. Recently there has been an interest in specifying a time series through a copula. The association of a copula to a Markov process dates back to Darswo and Olsen (1992), who states the necessary and sufficient conditions to specify a time series process based on a bivariate copula to be first-order Markov. We provide a summary of the vital results of Darswo and Olsen (1992).

Definition 4.2.1. *Let A and B be two copulas, and $u, v \in [0,1]$. The product $A * B$: $[0,1]^2 \mapsto [0,1]$, is given by*

$$(A * B)(u,v) = \int_0^1 \frac{\partial A(u,t)}{\partial t} . \frac{\partial B(t,v)}{\partial t} dt,$$

where $*$ denotes the product operation on copulas. Darswo and Olsen (1992) proofs the product $A * B$ to be a copula. Now before linking the copula to a Markov process, let us first define a stochastic process. A stochastic process is a collection of random variables $\{X_t\}_{t \in T}, T \subseteq \mathbf{R}$. t could be considered as index of time, and X_t as state of process at time t. For each $s, t \in T$, let F_s and F_t denote the respective continuous margins of X_s and X_t, and H_{st} be their joint distribution function. Similarly, let C_{st} be the copula of X_s and X_t respectively, then for all $x, y \in \mathbf{R}$,

$$H_{st}(x, y) = C_{st}(F_s(x), F_t(y)).$$

The process $\{X_t\}_{t \in T}$ is a Markov process if for every n, $t_1 < t_2 \dots t_n$ and $t \in T$ it satisfies

$$P(X_t \leq x | X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n) = P(X_t \leq x | X_{t_n} = x_n). \tag{4.2.1}$$

Let the conditional distribution (R.H.S.) in (4.2.1) be denoted as

$$P(s, x, t, y) = P(X_t \leq y | X_s = x).$$

Then we can present the vital theorem of Darswo and Olsen (1992), which provides the relationship of Chapman-Kolmogorov equations for a Markov process to the copulas of the random variables in the process.

Theorem 4.2.2. *Let $\{X_t\}_{t \in T}$, be a stochastic process, and let C_{st} denote the copula of random variables X_s and X_t, for each $s, t \in T$. The following are equivalent:*
1. The transition probabilities $P(s, x, t, A) = P(X_t \in A | X_s = x)$ of the process satisfy the Chapman-Kolmogorov equations,

$$P(s, x, t, A) = \int_{-\infty}^{\infty} P(u, \xi, t, A) P(s, x, u, d\xi),$$

for all Borel sets A, for all $s < t$ in T, for all $u \in (s, t) \cap T$ and for almost all $x \in \mathbf{R}$.
2. For all $s, u, t \in T$ satisfying $s < u < t$,

$$C_{st} = C_{su} * C_{ut}.$$

See Darswo and Olsen (1992) for the proof of Theorem 4.2.2. Nelson (2006) provides examples for constructing Markov processes using Theorem 4.2.2.

Theorem 4.2.2 only provides the necessary conditions, not the sufficient condition for a bivariate copula to produce a first-order Markov process. Darswo and Olsen (1992) also further provide the sufficient condition, but before presenting that, first, let us assume that A is an m dimensional copula CDF and B be a n dimensional copula CDF with both their support being on $[0, 1]^m$ and $[0, 1]^n$ respectively, then their product $A \star B$: $[0, 1]^{m+n-1} \mapsto [0, 1]$ ($\star$ denotes the product) is defined via

$$A \star B(x_1, \dots, x_{m+n-1}) = \int_0^{x_m} \frac{\partial A(x_1, \dots, x_{m-1}, \xi)}{\partial \xi} . \frac{\partial B(\xi, x_{m+1}, \dots, x_{m+n-1})}{\partial \xi} d\xi.$$

Then the theorem ensuring sufficient condition for a first-order Markov process through a copula can be stated.

Theorem 4.2.3. *A real valued stochastic process $\{X_t\}_{t\in T}$ is a Markov process if and only if for all positive integers n and for all $t_1, \ldots, t_n \in T$ satisfying $t_k < t_{k+1}, k = 1, \ldots, n-1$,*

$$C_{t_1,\ldots,t_n} = C_{t_1 t_2} \star C_{t_2 t_3} \star \cdots \star C_{t_{n-1} t_n},$$

where $C_{t_1,\ldots,t_n}$ is the copula of $X_{t_1}, \ldots X_{t_n}$ and $C_{t_k t_{k+1}}$ is the copula of X_{t_k} and $X_{t_{k+1}}$.

For prove of Theorem 4.2.3, see Darswo and Olsen (1992). We can then state, a process $\{X_t\}_{t=1}^{\infty}$ constructed through Theorem 4.2.3 is a stationary Markov process based on the copula C, or a C Copula based Markov chain. Lentzas and Ibragimov (2008) extends the framework of Darswo and Olsen (1992) for higher-order Markov processes.

The above is the most general specification of copula based Markov process, where the marginals distributions can vary over time, but Darswo and Olsen (1992) provide only the probabilistic properties for such specifications. To best understand the practicality of setting out a Markov process through a copula, let us assume the marginal distributions of the random variables do not vary over time, then Sklar (1959) theorem can be written to represent a first-order Markov chain $\{Y_t\}_{t=1}^{T}$ through copula C as

$$H(y_t, y_{t-1}) = C(F(y_t), F(y_{t-1})), \tag{4.2.2}$$

where H is the joint distribution, F is the marginal distribution (constant through time) and T is the length of the time series. Taking the partial derivatives w.r.t both y_t and y_{t-1} of (4.2.2) yields

$$\frac{\partial H^2(y_t, y_{t-1})}{\partial y_t \partial y_{t-1}} = h(y_t, y_{t-1}) = c(F(y_t), F(y_{t-1})).f(y_t).f(y_{t-1}), \tag{4.2.3}$$

where h is the joint density and c denotes the copula density. From (4.2.3) we can now state the conditional density of y_t given y_{t-1} as

$$h_{t|t-1}(y_t|y_{t-1}) = c(F(y_{t-1}), F(y_t)).f(y_t), \tag{4.2.4}$$

where $h_{t|t-1}$ denotes the conditional density. (4.2.4) is analogous to the previously mentioned Markov process, conditioning only upon the last observation. If F was assumed to be a normal distribution and C was specified as a Gaussian copula, then (4.2.4) would correspond to normal conditional density, hence an AR(1) process with normally distributed errors.

4.2.2 Copula Mixing Properties

For time series models generating a finite dimensional stationary Markov chain (like ARMA, GARCH etc.) it is vital to ask whether they satisfy weak dependence conditions such as geometric ergodicity. Similarly, for a copula-based time series model, we have to pose the same questions and see when such conditions will be satisfied.

It is crucial to understand that even though Markov processes generated through copula models with high tail dependence and commonly used lag numbers, behave like long memory time series, but in terms of mixing properties they are in fact weakly dependent and short memory. Given the non-linearities present in such models, powerful limiting theorems are required. Chen and Fan (2002) present conditions on copula-based Markov chains which are β-mixing with either exponential or polynomial decay rate, and for Gaussian and EFGM copulas geometric β-mixing is established. The conditions are independent of the invariant (marginal) distribution and only depend upon the copula specification.

Geometric ergodicity, which implies β-mixing is the strongest mixing property proven so far. Of course, β-mixing is a weaker assumption than φ-mixing and a covers a more general case of non-i.i.d series. We will briefly present vital results from Beare (2010a), but before that let is us state the definition of β-mixing.

Definition 4.2.4. *Let $\mathbf{Z} = \{Z_t\}_{t=-\infty}^{\infty}$ be a stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and for any $i, j \in \mathbb{Z} \cup \{-\infty, \infty\}$, let σ_i^j denote the σ-algebra generated by the random variables Z_k (where $s \leq k \leq t$). Then for any positive integer k, the β-mixing coefficient $\{\beta_k : k \in \mathbb{N}\}$ of the stochastic process $\mathbf{Z}$ is defined as*

$$\beta_k = \sup_n \mathop{\mathrm{E}}_{B \in \sigma_{-\infty}^n} \left[\sup_{A \in \sigma_{n+k}^{\infty}} \left| \Pr[A|B] - \Pr[A] \right| \right].$$

$\mathbf{Z}$ is said to be β-mixing if $\lim_{k\to\infty} \beta_k \to 0$, which is referred to as strong regularity. β_k measures the dependency between events taking place more than k units of time in the past. Definition 4.2.4 is taken from Yu (1994), but an equivalent to it was originally given in Volkonskii and Rozanov (1959). Beare (2010a) asserts a strong theorem over a copula C which if not-violated ensures geometric β-mixing. Before stating the theorem, we need to define Maximal correlation ρ_c of a copula C

$$\sup_{f,g} \left| \int_0^1 \int_0^1 f(x)g(y)C(dx, dy) \right|. \tag{4.2.5}$$

The supremum is taken over all $f, g \in L_2[0,1]$ such that $\int f = \int g = 0$ and $\int f^2 = \int g^2 = 1$. The integral in (4.2.5) is a Lebesgue-Stieltjes integral (see Rényi (1959) for details of maximal correlation). Then Beare's theorem is,

Theorem 4.2.5. *If C is symmetric and absolutely continuous with square-integrable density c, and that $\rho_c < 1$. Then there exists $A < \infty$ and $\gamma > 0$ such that $\beta_k \leq Ae^{-\gamma k}$ for all k.*

A copula is symmetric if $C(x, y) = C(y, x)$ for all $x, y, \in [0, 1]$, which implies the Markov chain Z_t is reversible. For the full proof of Theorem 4.2.5 see Beare (2010a). Some of the commonly used parametric copula families which satisfy Theorem 4.2.5 are Gaussian, Frank and Farlie-Gumbel-Morgenstren copulas. The pivotal requirement for geometric β-mixing in Theorem 4.2.5 is the square integrability of the density c and the symmetry, which Beare proves is not satisfied for copula families exhibiting upper or lower tail dependence, like Clayton, Gumbel and Student-t copulas. Hence Beare (2010a) is only able to establish β-mixing for asymmetric and tail dependent copulas. Chen et al. (2009) show processes generated through Clayton, Gumbel and Student-t copula to be all geometric ergodic. Beare (2010b) gives conditions for Archimedean copulas (of course including Clayton and Gumbel) through their generators, which ensure geometric ergodicity. The only assumption he imposes is such copula families stated in Nelsen (2007) be all regularly varying at the boundaries.

To summarize, for most of the copula families generally employed, geometric β-mixing has been established. Such properties are vital, as consistency and asymptotic normality for estimators of copula functions are conditional upon weak dependence in the time series considered.

4.2.3 Copula Estimation

Before introducing our semi-parametric approach to estimate the copula parameters using Bayesian techniques in Section 3 and 4, we provide a brief summary of various copula estimation techniques based on Maximum Likelihood (ML). These depend upon the assumptions made regarding the invariant distribution F, the copula C and the method of inference on the parameters. The estimation techniques for a first-order Markov chain through copulas is analogous to the bivariate contemporaneous dependence (generally seen through out the copula literature).

4.2.3.1 Full Parametric Approach

A fully parametric based estimation of the copula specification requires knowledge of the parametric marginal distribution and the parametric copula family. Let $\Gamma' = (\Lambda', \Theta')$ be the parameter vector which needs to be estimated. We can separate from Γ, the parameter vector Λ associated with the marginal distribution F (assuming F stays constant throughout the process), and Θ, the parameter vector corresponding to the copula C. In terms of the maximization problem of the probability density equation (4.2.3) over all T, can be stated as

$$\arg\max_{\Lambda,\Theta} \frac{1}{T}\left[\sum_{t=1}^{T} \log f(y_t; \Lambda) + \sum_{t=2}^{T} \log c(F(y_t; \Lambda), F(y_{t-1}; \Lambda); \Theta)\right]. \tag{4.2.6}$$

There are two commonly used method of inference for Λ and Θ. The first one is a standard estimator (MLE), where both Λ and Θ are estimated jointly. The second approach which is very widely employed in the literature is a 2-step approach known as Inference Functions for Margins (IFM), where first Λ is estimated, and conditional on $\widehat{\Lambda}$ (estimated) the copula parameter vector Θ is estimated in the second step. In practise IFM is easily implementable, as it implies obtaining u_t through $F(y_t)$, and simply plugging them into C. The IFM estimator like the MLE estimator is consistent and asymptotically normal under the usual regularity conditions. Computing the covariance matrix in the case of IFM is not straight forward both analytically and numerically, it requires methods like Jackknife and other such methods (see Joe (1997) for details). The asymptotic results obtained for these estimators in the context of cross-dependency are also preserved for copula-based time series according to Joe (1997).

4.2.3.2 Semi-Parametric Approach

Chen and Fan (2002) implement a semi-parametric estimation technique to specify the dynamics of a stationary Markov chain through a copula. They make no assumption regarding the invariant (or marginal) distribution and hence their estimation technique and inference is robust to misspecification of the marginal distribution. They propose a 2-step estimator, where the sample pseudo likelihood criterion does not depend upon the first-step estimator of the marginal distribution function. Then the estimation problem of Θ is

$$\arg\max_{\Theta} \tfrac{1}{T} \textstyle\sum_{t=2}^{T} \log c(\widetilde{F}(y_{t-1}), \widetilde{F}(y_t); \Theta),$$

where $\widetilde{F}$ denotes the Empirical CDF and the marginal specification is

$$\widetilde{U}_t = \widetilde{F}(Y_t) = \tfrac{1}{T+1} \textstyle\sum_{s=1}^{T} \mathbb{1}(Y_s \leq Y_t).$$

$\widetilde{F}$ is the re-scaled empirical distribution function, although the kernel smoothed estimator could also be used. Chen and Fan (2002) establish the consistency and asymptotic normality of the above semi-parametric estimator. They verify these properties for Markov processes generated through Gaussian, Clayton and Frank copulas.

Chen et al. (2009) propose a Sieve ML Estimation for copula-based time series. They show that the sieve MLE of any smooth function is root-n consistent, asymptotically normal and efficient, and the sieve likelihood ratio statistics is chi-squared distributed. They perform Monte Carlo studies to show that their technique has smaller bias and variance compared to the two-step estimator of Chen and Fan (2002) for series generated by Clayton, Gumbel and other copulas exhibiting tail dependence.

4.2.4 Non-Parametric Approach

A completely non-parametric specification is also feasible, where both the marginal distribution and the copula are empirically estimated. The non-parametric estimator of C could be defined as $\widetilde{C}$

$$\widetilde{C}(u_{t-1}, u_t) = \widetilde{H}(\widetilde{F}^{-1}(u_{t-1}), \widetilde{F}^{-1}(u_t)), \tag{4.2.7}$$

where $\widetilde{F}^{-1}$ is the non-parametric estimator of the pseudo-inverse F^{-1}. Here H is the empirical distribution function $\widetilde{H}(y) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(Y_{t-1} \leq y, Y_t \leq y)$. Doukhan et al. (2009) establish asymptotic normality of the smoothed copula process with kernel estimates for H and the univariate margin F in (4.2.7) for weakly dependent vector-values sequences. The literature on applications of such techniques is limited, see Fermanian and Scaillet (2003) for details.

4.2.5 Discrete Marginals

Although the copula framework allows us to separate out the marginal behaviour from the dependency given by a copula C, the type of random variable (continuous or discrete) under consideration will have an effect on C. We present the argument from Genest and Nešlehová (2007), where they discuss the problems faced with count data. C is uniquely defined and obtained as in (4.2.2), only if H is continuous. In the case it is discrete there are several functions A such that

$$H(y_t, y_{t-1}) = A(F(y_t), F(y_{t-1})),$$

We can obtain a solution from above as

$$B(u_t, u_{t-1}) = H(F^{-1}(u_t), F^{-1}(u_{t-1})).$$

We are able to solve for B of course, but this does not imply B is a copula function or even a distribution function. This is referred to as the identifiability issue, as it requires identifying set of copulas for which A can replaced by $C \in \mathbf{C}_H$ (i.e. some set of copulas $\mathbf{C}_H$), for which (4.2.2) holds. For such set of copulas we have to understand what are its smallest and largest possible elements. Generally a copula C for any $u, v \in [0,1]$ is well defined over the Fréchet-Hoeffding bounds. The bounds for a copula C are given by

$$W(u,v) = \max(u+v-1, 0) \leq C(u,v) \leq \min(u,v) = M(u,v),$$

where W and M are the Fréchet-Hoeffding lower and upper bounds. But in the case of discrete margins sharper bounds are required

$$C_H^-(u,v) \leq C(u,v) \leq C_H^+(u,v),$$

which apply to any $C \in \mathbf{C}_H$. Such bounds are proven to exist by Carley (2002). Genest and Nešlehová (2007) show that the class $\mathbf{C}_H$ is quite large, and hence the problem of unidentifiability for such margins is to be considered.

From an empirical perspective, perhaps the problem is best understood by interpreting θ (copula dependency parameter) as a function of Kendall's tau (concordance measure), which as Genest and Nešlehová (2007) mention can be computed from the ranks of the observed series. Compared to Spearman's rho measurement of association, Kendall's tau has simpler formulas for copula families (see Nelson (1999)). Such concordance measure provide meaning full interpretation for the copula parameters, as they are bounded on the interval $[-1, 1]$.

Given our observed time series is $(Y_1, Y_2, \ldots, Y_T)$, let the corresponding ranked series be denoted as $(K_1, K_2, \ldots, K_T)$. As we are interested in a first-order Markov chain of order one, so the number of pairs are $(T-1)$ for a bivariate copula. Assuming the mapping of Kendall's tau τ from the copula C ($C \in C_\theta$, parametric set) parameter $\tau : \theta \mapsto \tau(C_\theta)$ is one-to-one, let $\widetilde{\theta}_T$ be a method-of-moment estimator of θ, such that $\tau(C_\theta) = \tau_T$, where τ_T is the sample T value of the Kendall's tau measure equal to

$$\tau_T = 1 - 2N/\tbinom{T-1}{2},$$

where N is the number of discordances given as

$$N = \textstyle\sum_{i=1}^{T-1} \sum_{j=1}^{T-1} \mathbb{1}(K_i < K_j, K_{i+1} > K_{j+1}).$$

If the choosen copula is a Clayton copula, then the corresponding $\widetilde{\theta}_T$ would be

$$\widetilde{\theta}_T = 2\frac{\tau_T}{1-\tau_T}.$$

Now let us understand what happens to such an estimator, given either continuous or discrete marginal distribution. In the case of the former the estimator $\widetilde{\theta}_T$ of θ is consistent and asymptotically normal and the estimate of τ_T based on either the observed $\{Y\}_{t=1}^T$ or $\{U\}_{t=1}^T$, is equal to each other, but for discrete data the equality might not hold. The quantile functions in a discrete case, are not strictly monotone, and $P[(Y_i = Y_j, Y_{i+1} = Y_{j+1})] \neq 0$ for some $i \neq j$. The mapping of $\{Y\}_{t=1}^T$ to $\{U\}_{t=1}^T$ through the marginal distribution, is not one-to-one. Regardless of the size of the time-series, this discretization is irreversible. We could introduce some commonly employed solutions for such a problem, like splitting the ties, ignoring them or adjusting them. But the mere fact that ties could occur, bias the estimates of θ, hence rank based methods are not appropriate for discrete margins (see Genest and Nešlehová (2007) for details).

To summarize, in discrete data case, concordance type measure are not independent on the functional form of the marginal distributions. The Kendall's tau measure reduces in the presence of ties. Trivedi and Zimmer (2006) mention maximization of likelihood with discrete margins poses computational difficulties and proposes to perform continuation transformation, where each discrete margin is made continuous by adding some noise (Uniform $[0, 1]$ draw), then proceed with copula estimation, with continuous margins. Although such a process would imply misspecification of the margins. Hoff (2007) calls marginal parameters as nuisance parameters, especially for discrete data, and derives a likelihood which treats the copula arguments as latent variables and relies on the fact that they have the same order statistics, as the observed data. Other similar Bayesian methods are specified in Pitt et al. (2006) and Smith and Khaled (2012) to deal with discrete margins. Generally the interpretation of θ (copula parameter) does not have the same meaning in case of discrete margins as it has for continuous margins.

§ 4.3 Framework

Before specifying the Bayesian sampling scheme, we introduce the necessary framework regarding the order statistic and the association of each instance of $\mathbf{U}$ within the Markov chain.

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_T)$ be a strictly stationary time series originating from an unknown marginal distribution F, and $\mathbf{U} = (U_1, U_2, \ldots, U_T)$ be the series of uniforms, with each instance being in $[0, 1]$. Let the Markov chain be generated through copula C, then we can specify the Data Generating Process (DGP) for a first-order Markov chain as

$$c(u_t|u_{t-1}; \Theta), \text{ for } t = 2, \ldots, T,$$
$$y_t = j, \text{ if } \max\Big\{u_s;\ F : j-1 \mapsto u_s\Big\} < u_t < \min\Big\{u_s;\ F : j+1 \mapsto u_s\Big\},$$
$$\text{where } j \in J \text{ (discrete outcomes).}$$

Where Θ is the parameter vector associated to C, and j is a discrete observation belonging to set of possible values in J. Each $u_t \in [0, 1]$ is generated through the conditional copula density c, and the corresponding y_t determined through the maximum and minimum of the uniforms corresponding to the neighbouring order statistics of j, as seen in Figure 4.1. The DGP described above is set out for a discretely-varied time series, in case we are dealing with a time series of continuous type random variables, the correspondence of u_t to y_t is one-to-one.

If F is known, and belongs to either a parametric family or non-parametric (empirical distribution), then one of the estimators stated in Section 2.4 could be employed for Θ. For a continuous margin such estimators would yield constant and asymptotically normal estimates of Θ, but in case F is a discrete distribution, ML methods can fail with convergence of the likelihood and results will be biased from using continuous transformation.

We treat F as completely unknown, and hence $\mathbf{U}$ is unobtainable and considered as a series of latent variables. The only available information available related to F is that it is a non-decreasing monotonic function, and could either be a continuous or a discrete distribution. In case the margin is continuous, F^{-1} will be a one-to-one mapping function, and for a discrete margin a many-to-one function. It is the first time a

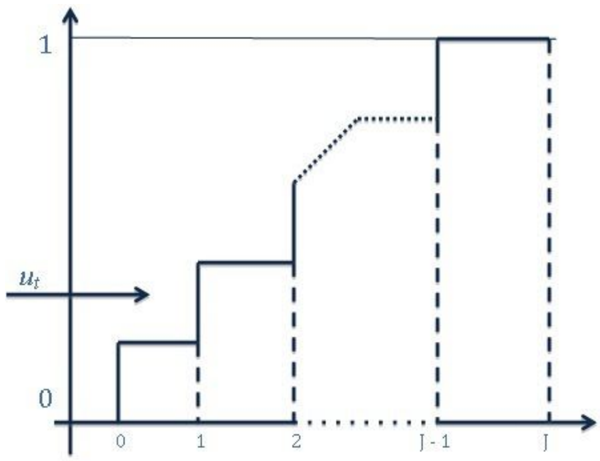


Figure 4.1: *Mapping of generated u_t*

completely non-parametric specification has been assumed for the marginal distribution using for Markov chain type time series framework, both for continuous and discrete type outcomes. This overcomes any form of marginal misspecification, and allows to combine any type of marginal behaviour with non-gaussian type of temporal dependence.

Given that F is non-decreasing, we know the order statistics of the uniforms generated through the unknown F will be dictated by the order statistics of the observed $\mathbf{Y}$, and this is the only information known with certainty. But there is still the uncertainty of the actual value of $\mathbf{U}$, and the degree of uncertainty depends upon the discrete data (low count implying more uncertainty). In case we have a time series of binary outcomes, there is really only two ranks, and hence we have more uncertainty. We can provide a formal definition for the order statistics of the time series.

Definition 4.3.1. *Let the rank of the observation at time t, y_t, be denoted as k_t. Hence $y_t = y^{(k_t)}$, and for each t*

$$y^{(k_t-1)} < y_t < y^{(k_t+1)}, \text{ and,}$$
$$u^{(k_t-1)} < u_t < u^{(k_t+1)}.$$

$y^{(k_t)}$ is the order statistic of y_t. u_t has the same rank k_t, as y_t.

Definition 4.3.1 simply states that given that F is non-decreasing and monotonic, the unobserved $\mathbf{U}$ have to obey the same order statistics as that of $\mathbf{Y}$. We keep the time

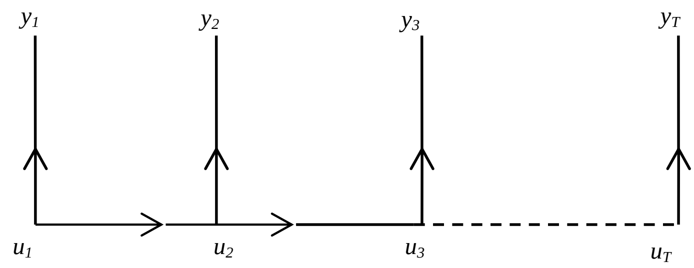


Figure 4.2: *DAG of Latent Variable*

indexing on the ranks, as unlike cross-sectional analysis, the time stamp on an observation is important. Note, we have strict-inequality for the ranks of the observed data, implying any ties are left unresolved.

As we are capturing the temporal dependence of the series $\mathbf{U}$, each instance u_t is related to its neighbor in time and the corresponding order statistic through $\mathbf{Y}$. This is perhaps best understood by employing a Directed Acyclic Graph (DAG) in Figure 4.2, where we see that $\mathbf{U}$ is a Markov chain, and the observation y_t are independent of each other conditional upon u_t. Starting from u_1 the chain moves forward till the last value u_T. We see how each u_t is connected to its neighbors in time (u_{t-1} and u_{t+1}) and the corresponding y_t, from which the only information retrieved is k_t, the rank.

Using the above framework of the DAG, we can now model the structure of the Markov chain in a Bayesian setup.

§ 4.4 Bayesian Sampling Scheme

We specify a general sampling scheme to estimate the copula parameters, which will capture the temporal dependence within a stationary Markov chain. The estimation technique is general to any copula family and makes no assumption regarding the marginal distribution, and hence can accommodate both continuous and discrete type margins. The sampling scheme can be separated into two different stages, the first stage involves sampling $\mathbf{U}$ conditional upon the copula parameter Θ, and in the second stage we draw

Θ conditional upon $\mathbf{U}$. This implies all the uniforms are considered as auxiliary variables and have to be sampled.

Before proceeding with the sampling scheme, let us define the necessary notation. Through Bayes theorem, let the posterior of Θ be given as

$$p(\Theta|\mathbf{U}) \propto p(\Theta) \times p(\mathbf{U}|\mathbf{Y};\Theta), \tag{4.4.1}$$

where π denotes the posterior of Θ. To make our scheme general across various copula families, we need to re-parametrize the copula parameter vector Θ. Copula families support different ranges, for a Gaussian and Student-t copula the dependence parameter lies in $[-1, 1]$. Whereas for most Archimedean copulas the upper or lower bounds are defined up to infinity, like the range of Clayton copula parameter is $(0, \infty)$. We can transform them all to be defined over the real line $\mathbb{R}$. Let the mapping be $\mathbf{Z}(\Theta) = \Psi$, where $\Psi \in \mathbb{R}$ and $\mathbf{Z}$ (See Appendix A.1 for various transformations) represents a vector of functions, which has the same dimension as Θ. This re-parametrization will change the posterior defined in (4.4.1). The prior distribution $p(\theta)$ has to be transformed over to prior of Ψ, and we have to consider the Jacobian matrix associated to such a reformulation. The prior $p(\Psi)$ for Ψ will then be defined as

$$p(\Psi) \propto p(\Theta)\left|\frac{\partial\Theta}{\partial\Psi}\right|,$$

where $\left|\frac{\partial\Theta}{\partial\Psi}\right|$ is the determinant of the Jacobian matrix. Regardless of which copula is now chosen, we have a support over the real line for the parameters associated. The posterior defined in (4.4.1) now becomes

$$p(\Psi|\mathbf{U}) \propto p(\Psi) \times p(\mathbf{U}|\mathbf{Y};\Psi).$$

Finally, we can now proceed with specifying the two stage sampling scheme, where first we sample from $p(\mathbf{U}|\mathbf{Y};\Psi)$, followed by sampling from $p(\Psi|\mathbf{U})$.

4.4.1 Sampling from $p(\mathbf{U}|\mathbf{Y};\Psi)$

We see from Figure 4.2 how each instance of $\mathbf{U}$ is linked by its neighbours in time and the corresponding order statistic from $\mathbf{Y}$. Assuming a first-order Markov chain, we can write the conditional probability for each u_t, where $1 < t < T$ as

$$p(u_t|\mathbf{U}_{\backslash \mathbf{t}}, \mathbf{Y}; \Psi) = p(u_t|u_{t-1}, u_{t+1}, y_t; \Psi), \tag{4.4.2}$$

where $\mathbf{U}_{\backslash \mathbf{t}}$ is the complete series $\mathbf{U}$ without u_t. Given that we have a Markov chain of order one and using the information from the DAG, the conditioning of $\mathbf{U}_{\backslash \mathbf{t}}$ can be reduced to u_t's connected neighbours in time (i.e. u_{t-1} and u_{t+1}). As we mentioned previously, the only information available from conditioning u_t on y_t is, if $y^{(k_t)}$ is the order statistic of y_t then k_t is also the rank of u_t. This implies u_t has to lie between $u^{(k_t-1)} < u_t < u^{(k_t+1)}$ to maintain the order statistics, and the size of the interval depends upon the degree on discreteness. So we can simplify (4.4.2) as

$$p(u_t|u_{t-1}, u_{t+1}, u^{(k_t-1)}, u^{(k_t+1)}; \Psi) = p(u_t|u_{t-1}, u_{t+1}; \Psi)\mathcal{I}(u^{(k_t-1)} < u_t < u^{(k_t+1)}). \tag{4.4.3}$$

We cannot directly sample from (4.4.3) through a bivariate copula density (a first-order Markov chain corresponding to bivariate copula), but applying Bayes theorem further, we can write $p(u_t|u_{t-1}, u_{t+1}; \Psi)$ as

$$p(u_t|u_{t-1}, u_{t+1}; \Psi) \propto p(u_t|u_{t-1}; \Psi) \times p(u_{t+1}|u_t; \Psi).$$

Now we have two conditional distributions and we introduce the conditional copula, let C be the copula distribution, and let $c_{t|t-1}$ be the bivariate conditional copula density of u_t given u_{t-1}, corresponding to C (see Appendix A.1.1 for various copulas density formulation). We model the conditional probability of u_t through the conditional copula density as,

$$p(u_t|u_{t-1}; \Psi) = c_{t|t-1}(u_t|u_{t-1}; \Psi).$$

Hence the copula represents the transition density of the Markov chain. The same holds for conditioning of u_{t+1} to u_t. Now we can easily sample from the conditional distribution $C_{t|t-1}(u_t|u_{t-1}; \Psi)$, and evaluate the draw through Metropolis-Hasting (M-H) algorithm using $c_{t+1|t}(u_{t+1}|u_t; \Psi)$. The sampling scheme is given as

for each $u_t, (t = 1, \ldots, T)$,

compute $u^{(k_t-1)}$ and $u^{(k_t+1)}$, given $(y^{(k_t-1)} < y_t < y^{(k_t+1)})$,

$$\text{sample } u_t^* \text{from} C_{t|t-1}(u_t|u_{t-1};\Psi)\mathcal{I}(u^{(k_t-1)} < u_t < u^{(k_t+1)}),$$

$$\text{compute } \alpha_u = \min\left\{1, \frac{c_{t+1|t}(u_{t+1}|u_t^*;\Psi)}{c_{t+1|t}(u_{t+1}|u_t;\Psi)}\right\}.$$

Sampling from $C_{t|t-1}$ is easier than $c_{t|t-1}$, as most copulas have a closed inverse form for the conditional distribution. The above scheme is repeated for all u_t. The truncated intervals are updated if the drawn u_t^* is accepted. The sampling is performed in the order of the $\mathbf{U}$ dictated by the time order, but the intervals have to be maintained regardless of time. Through this scheme we obtain an updated sample of $\mathbf{U}$. For large enough T (for continuous data) and high count data (discrete data), the interval $(u^{(k_t-1)}, u^{(k_t+1)})$ becomes smaller and the acceptance probability α_u gets close to one. In fact anything uniformly sampled through the interval can be accepted, and we would not need to pass them through the M-H step to evaluate the conditional copula of u_{t+1}. Missing values can also be generated through this scheme. We could also consider higher-order Markov chain, for example for a second-order Markov chain, we will require a trivariate copula and proceed with similar sampling procedure.

4.4.2 Sampling from the Posterior $p(\Psi|\mathbf{U})$

Now we can proceed with sampling from the posterior of Ψ. Unlike the Gaussian copula (see Hoff (2007)), most copula families do not have the full conditional available to sample from, and a Markov Chain Monte Carlo (MCMC) based on M-H algorithm has to be adopted.

Within the M-H framework, we have to choose an adequate proposal distribution $g(\Psi)$. A multivariate t-distribution with the mean equal to the mode of the posterior, and the variance equal to the negative inverse of the information matrix computed at the mode, can be used as a proposal distribution. Such a Laplace-type proposal has been used in the literature (see Chib and Greenberg (1998), Chib and Winkelmann (2001) and Pitt et al. (2006)). A multivariate t-distribution with ν degrees-of-freedom as opposed to a Normal distribution is preferred, as it would dominate in the tails of the true density. The advantage of such a proposal density is we do not need to consider tuning of parameters, to attain some acceptance probability.

We choose a flat prior for Θ, $p(\Theta) = 1$. Hence the re-parametrized Ψ's prior will be $p(\Psi) \propto \left|\frac{\partial\Theta}{\partial\Psi}\right|$. To use a Laplace approximation, we need to employ a Maximum a Posterior Probability (MAP).

$$\widehat{\Psi}_{MAP}(\mathbf{U}) = \arg\max_{\Psi} p(\mathbf{U}|\mathbf{Y};\Psi)p(\Psi),$$

where $\widehat{\Psi}$ denotes the estimated mode of Ψ. The log of the posterior can then be written as

$$\log p(\Psi|\mathbf{U}) \approx \log g(\Psi|\widehat{\Psi};V).$$

$V = -I^{-1}$ and $I = \left[\frac{\partial^2 p(\mathbf{U}|\mathbf{Y};\Psi)p(\Psi)}{\partial\Psi\partial\Psi'}\right]_{\Psi=\widehat{\Psi}}$ is the information matrix evaluated at the mode. We can finally draw from

$$\Psi \simeq t_\nu(\widehat{\Psi}, V),$$

let the drawn value be denoted as Ψ^* and the current value be Ψ, then Ψ^* can be evaluated using M-H with the following acceptance probability

$$\alpha_\Psi = \min\left\{1, \frac{\pi(\Psi^*)t_\nu(\Psi|\widehat{\Psi},V)}{\pi(\Psi)t_\nu(\Psi^*|\widehat{\Psi},V)}\right\},$$

where $t_\nu(\Psi|\widehat{\Psi},V)$ denotes the density of $t_\nu(\widehat{\Psi},V)$ evaluated at Ψ. The acceptance ratio is very high, as the proposal density is close to the true density. We arbitrarily choose $\nu = 6$ to dominate in the tails. A new mode $\widehat{\Psi}$ is found at each iteration, conditional on the new updated $\mathbf{U}$ sample. To avoid computational burden of finding a new mode at each iteration, we could update the mode every hundredth time or so, but we cannot not change it as the $\mathbf{U}$ get sampled at each iteration. Numerical techniques such as Newton-Raphson are required to locate the mode, and they are found within few steps of the algorithm search. Re-estimating the mode at each iteration could become computationally intensive, and could be avoided by updating it at regular intervals.

§ 4.5 Alternative Models

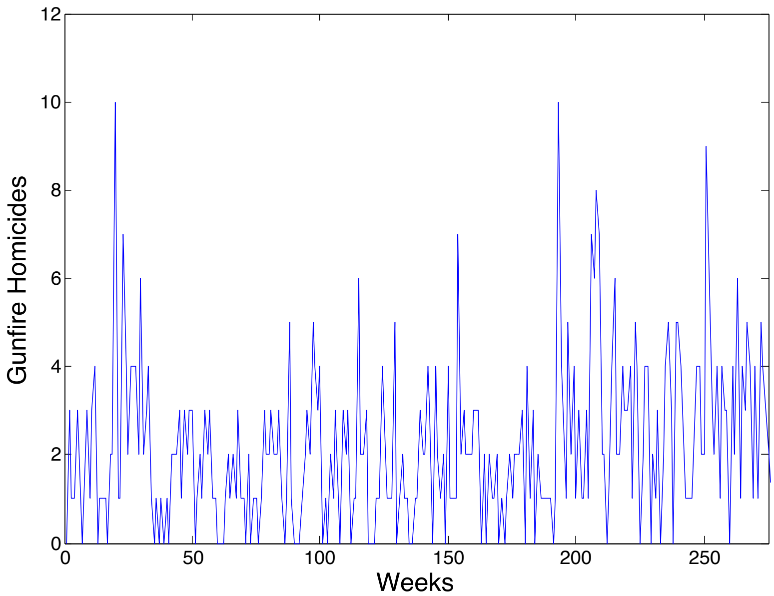
We have proposed a novel approach to specify the dynamics of a discretely-varied time series, where no assumption whatsoever is made on the marginal distribution and the transition density is given through a parametric copula allowing complex dependence patterns to be captured. However, this is not the only technique available to a practitioner for modelling time series of non-continuous type. The most commonly used and employed model for count data is the Integer Valued Autoregressive Process (INAR) of Al-Osh and Alzaid (1987), it is akin to an AR model, but the error term is appropriately assumed to be of discrete type (binomial, poison distribution etc.). Such model has been successfully employed by Freeland and McCabe (2004) for capturing the dynamics of weekly wage loss claims data in Province of British Columbia, Canada. There is also a class of models, to which our technique is much similar to is that of Pitt and Walker (2001), where a time series is analysed through a Markov chain allowing for the marginal distribution to be specified independent of the transition density. They, similar to us (next section) model the dynamics of Firearm Homicides in Cape Town, South Africa by assuming the marginals to be poisson and the transition given through a gamma process.

It is absolutely vital to compare these models to our technique in terms of good-of-fit and in and out-of-sample prediction, as ultimately time-series models are used for forecasting. Now after we have successfully estimated the copula parameter ($\widehat{\Theta}$, posterior mean) capturing the temporal dependence, we want to to be able to conduct inference on the observed $\mathbf{Y}$, like point predictions or computing confidence intervals etc. As discussed in detail, the copula is estimated on the latent uniform variables $\mathbf{U}$, and the mapping from the observed $\mathbf{Y}$ to the $\mathbf{U}$ is done through the marginal distribution F, which in our case is completely unknown. Independent of the copula model, we cannot determine this mapping. Our aim in keeping the marginal as completely unknown, was to tackle the problems associated with discrete data outcomes and to avoid any sort of misspecification. If we were to assume a parametric form for F after estimating the copula parameter and proceed with conducting inference on $\mathbf{Y}$, that would simply defeat the purpose of our technique, which is purely non-parametric in terms of the margins and would not offer a reasonable comparison with the other models. Also assuming let us say an empirical distribution in case of count data outcomes, would induce bias due to the ties observed in the ranks. In someway this is a short-coming of our approach and currently our on going research is to be able to conduct inference in a way which would coincide with how we estimated the copula parameter, that is through the information available only within the order statistics.

§ 4.6 Data Example (Firearm Homicides)

We now present a real data application, where there could be persistence within a discretely-varied time series. To be specific, we are looking at the weekly firearm homicides in Cape Town, South Africa from the period of 1st January 1986 up to 31st February 1991 (275 observations)[1], as given in Figure 4.3.

Figure 4.3: *Firearm Homicides South Africa*



Generally the period of the data sample corresponds to a time when areas in and

[1]Data obtained from MacDonald and Zucchini (1997).

around Cape Town experienced rapid urbanization. Through such development local gangs form in these areas, and clustering of high count of homicides could be associated with it. Cape Town also attracted people from different regions of South Africa to settle in, and differences in social norms could be another contributing reason.

We could employ various copula models to capture some specific form of temporal dependence present in such a series. Two copula families will be employed here, first a Gumbel copula which is a one-parameter copula exhibiting greater dependence in the right tail. Secondly, to emphasis that our method is general across copula families a standard Gaussian copula with no tail dependence parameter will be used, for the transition density.

The MCMC scheme is performed for 50,000 iterations. To deal with posterior correlation, we perform thinning by saving every 30th iterate for posterior analysis, also the first 5000 iterations are discarded for burn-in. Figure 4.4 presents the trace plot for both the Gumbel and the Gaussian copula, where we see the chain mixes well. We also present the autocorrelations plot of the final posterior sample for both of the used copulas in figure 4.5. The autocorrelation is lower than 0.02 after the 4th and 3rd lag for the Gumbel and the Gaussian copula parameter respectively. Further to ensure convergence to the stationary distribution, we performed multiple runs from different initial values. The reason for finding such high correlation, is due to the sampling of each u_t at a time, and then sampling Ψ based on the whole sample of $\mathbf{U}$. In a time series framework it is difficult to overcome such posterior correlation.

Apart from obtaining the posterior mean of Θ, we also compute the Kendall's Tau rank correlation coefficient for the copulas, which provides a meaningful interpretation of the dependence as it is defined over $[-1, 1]$. Computed as

$$\tau = 4 \iint_{[0,1]^2} C(u,v) dC(u,v) - 1. \tag{4.6.1}$$

For most copula families, a simple analytical solution for (4.6.1) exists, else it has to be numerically computed. From using both of the copulas, we see there is evidently some temporal dependence present within this time series, suggesting persistence. The Kendall's tau measure through both copulas is the same. For the Gaussian copula the tail

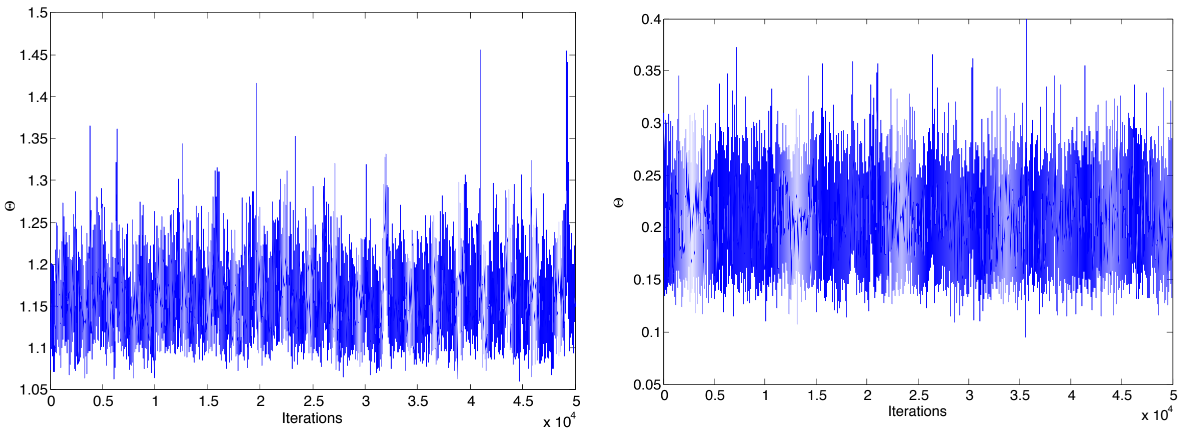


Figure 4.4: *Trace Plots: Gumbel & Gaussian Copula*

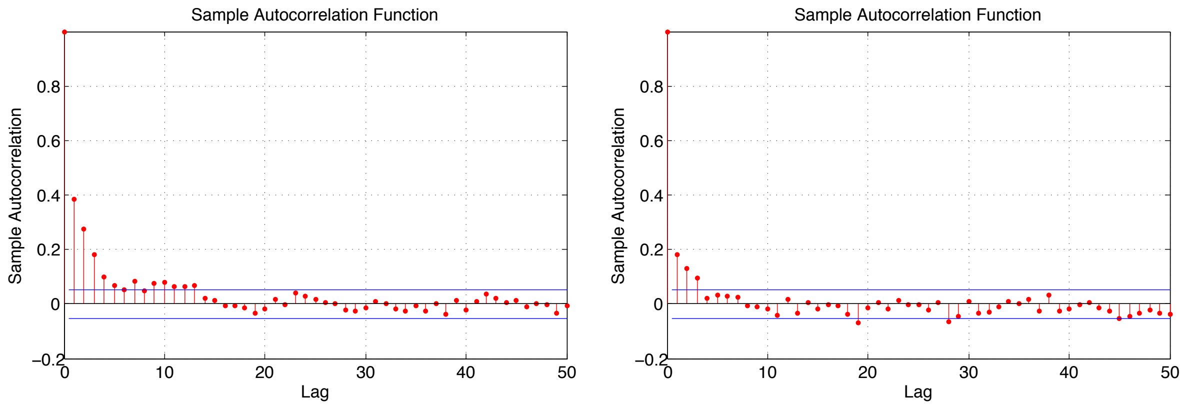


Figure 4.5: *Autocorrelation Plots: Gumbel & Gaussian Copula*

Table 4.1: *Posterior Distribution Inference (Firearm Homicides)*

Copula C	$E(\pi(\Theta))$ (Posterior mean)	τ (Kendall's tau)	λ_U (Upper tail dep.)
Gumbel	1.278	0.138	0.182
	(0.069)	(0.026)	(0.033)
Gaussian	0.209	0.134	-
	(0.031)	(0.020)	

dependence (lower and upper) parameter is 0, but for the Gumbel copula we can compute the upper tail dependence parameter through,

$$\lambda_U = \lim_{1^-} \frac{1 - 2u + C(u,u)}{u},$$

where $\lim_{1^-}$ is the limit approaching from the left. From Table 4.1 we can see there is significant upper tail dependence, which implies, through Gumbel copula there seems to be higher probability of observing high homicides counts this week if in the previous week there were high number of homicides.

Apart from the persistence observed within the time series, there could be other explanatory variables to determine the current period firearm homicides, like number of police deployment in local boroughs of Cape town over time, policy implementation, number of people moving to Cape Town from rural areas around and others. Such covariates can be considered either within copula framework where we could choose separate copula families to present the contemporaneous dependence or if our interest is not on these covariates but solely on the temporal dependence, we can filter out the unexplained variation in homicides through a regression where the covariates are considered.

§ 4.7 Further Discussion

Our proposed methodology completely requires the time series to be strictly stationary, as the method falls in the class in Markov based time series models, which require the joint probability distribution not to change when shifted in time. To be able to model the temporal dependence within a non-stationary times, we would first have to make it stationary by standard approaches (i.e. differencing etc), and then apply by method. By doing so of course the interpretation of the temporal dependence will be based upon the change in the observed values.

Another issue generally faced with such socio-economic data like firearm homicides is the possibility of observing structural breaks maybe due to policy implementation. We could perform standard structural break tests like Chow test, and if present compute the temporal dependence measure over appropriate sub-samples.

Generally in time series an econometrician is interested in making predictions of the variable of interest. In our setup, we cannot easily make predictions on the observed $\mathbf{Y}$ for two reasons, first, we make no assumption on the marginal distribution and it is completely unknown, which implies we cannot apply the inverse CDF to obtain the observed series. Secondly, the $\mathbf{U}$ are latent variables which change at each MCMC iteration and are filtered through the copula density, so cannot be used to transform back to the observed. We could assume an empirical CDF on the observed series, through which we can obtain predictions conditional upon the posterior estimates, but that would not fair well with the novelty of our approach.

In time series framework along with lags of the dependent variables, covariates can also also explain the endogenous variable. In our modelling scheme, we can have y_t conditioned on the covariates z_t' either through the same copula as y_{t-1} on y_t, or through a separate copula. If our interest is not on how covariates explain the endogenous variable but only to have an unbiased estimate of the temporal dependence measure or if the size of z_t' is very large, then we could also filter out the effects of the covariates through a separate regression before proceeding with the copula-based Markov chain.

§ 4.8 Conclusion

Time series models fail to be flexible enough to capture complex temporal dependence for data of different types (binary, count, ordered etc.). The entanglement of the marginal distribution with the conditional (transition) distribution, confine most models to specific problems. Models constructed for data of continuous type cannot normally be applied to discrete data type. Copula models are generally used to capture contemporaneous dependence, but recently they have been considered to describe a time series process (see Joe (1997), Chen and Fan (2002) and Chen et al. (2009) among others). Lentzas and Ibragimov (2008) show a copula-based Markov chain can act like long memory process. Joe (1997) proposes a parametric copula-based Markov chain for continuous data, where both the marginal distribution and the copula are parametrically specified. Chen and Fan (2002) present a semi-parametric copula-based Markov chain, where the marginal distribution is computed through an empirical distribution and employing Maximum Likelihood for the copula parameters, yields consistent and asymptotically normal results. However, for discrete data types such methods can create computational problems (algorithm failing to converge) and induce bias.

Hoff (2007) proposes a technique for cross-sectional data of mixed type (continuous-discrete), where the marginals are left completely unspecified and based only on the information contained in the order statistics a multivariate Gaussian copula is estimated. We extended Hoff's technique in a time series framework and make it general across various copula families. Such a technique can accommodate time series of both discretely and continuously varying random variables. A Bayesian sampling scheme is proposed, where we first sample the latent uniform variables conditional upon the copula parameters, and then the copula parameter conditional upon the sampled uniform series. To make the methodology general across copula families, a Laplace-type proposal for the posterior is presented, where each draw is evaluated through a Metropolis-Hasting algorithm.

We use a real data application based on the weekly count of firearm homicides in Cape Town, South Africa. Employing both a Gumbel and a Gaussian copula separately, we capture the significant persistence present within such a time series. Various quantities of interest, like tail dependence can also be computed through copulas. The MCMC technique works well. Appropriate thinning and discarding of the initial posterior draws is performed to ensure no autocorrelation in the final posterior sample.

We presented a case for first-order Markov chain to keep the intuition clear. A higher order Markov chain can easily be considered (see Ibragimov (2009)). Currently, we are working on model selection in terms of which copula best fits a time series, through computing the Marginal Likelihood. We are also performing diagnostics through Probability Integral Transformation (PIT), and comparing the model with other commonly used count time series models like Poisson Integer-valued Autoregressive Model (PoINAR), in terms of goodness-of-fit.

Part I

Annexes

- Appendix A -

Annexes to Chapter 3

§ A.1 Copula Families and Conditional Distribution

A.1.1 Copula Transformations

We can obtain analytical formulas for important copula properties through the general distribution function C. For the copula density c given the distribution function is twice differentiable

$$c(u, v|\Theta) = \frac{\partial^2 C(u, v|\Theta)}{\partial u \partial v}.$$

And the conditional distribution of a copula given as

$$C_{u|v}(u|v; \Theta) = \frac{\partial C(u, v|\Theta)}{\partial v}.$$

The conditional copula density $c_{u|v}(u|v; \Theta) = c(u, v|\Theta)$.

A.1.2 Clayton Copula

The Clayton copula is an Archimedean copula exhibiting strong joint left tail dependence. It is a one-parameter based copula. The various transformations for Clayton copula are given as

$$C(u, v|\alpha) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}),$$

$$c(u, v|\alpha) = (1 + \alpha)[uv]^{-\alpha-1}(u^{-\alpha} + v^{-\alpha} - 1)^{-2-1/\alpha},$$

$$c_{u|v}(u|v; \alpha) = v^{-\alpha-1}(u^{-\alpha} + v^{-\alpha} - 1)^{-1-1/\alpha},$$

$$C^{-1}_{u|v}(u|v; \theta) = \left[(uv^{\theta+1})^{-\theta/(1+\theta)} + 1 - v^{-\theta}\right]^{-1/\theta},$$

where $u, v \in [0, 1]$, and the copula parameter $\alpha \in [0, \infty]$.

A.1.3 Gumbel Copula

Gumbel copula is an Archimedean copula exhibiting strong joint right tail dependence. It is a one-parameter based copula. The various transformations for Gumbel copula are given as

$$C(u, v|\alpha) = \exp\left[- (\widetilde{u}^{\alpha} + \widetilde{v}^{\alpha})^{1/\alpha}\right],$$

$$c(u, v|\alpha) = C(u, v|\alpha)(uv)^{-1}\frac{(\widetilde{u}.\widetilde{v})^{\alpha-1}}{(\widetilde{u}^{\alpha} + \widetilde{v}^{\alpha})^{2-(1/\alpha)}}\left[(\widetilde{u}^{\alpha} + \widetilde{v}^{\alpha})^{1/\alpha} + \alpha - 1\right],$$

$$C_{u|v}(u|v; \alpha) = v^{-1}\exp\left[- (\widetilde{u}^{\alpha} + \widetilde{v}^{\alpha})^{1/\alpha}\right]\left[1 + (\frac{\widetilde{u}}{\widetilde{v}})^{\alpha}\right]^{-1+(1/\alpha)},$$

where $\widetilde{u} = -\text{Log}\, u$ and $\widetilde{v} = -\text{Log}\, v$, $u, v \in [0, 1]$, and the copula paramater $\alpha \in [0, \infty]$. The inverse of the conditional gumbel distribution can not be solved analytically and hence we have to rely on a numerical method.

A.1.4 Gaussian Copula

Gaussian copula is an Elliptical copula, which is completely symmetric and has zero probability for any left/right extreme dependence. It is a one-parameter based copula. The various transformations for Gaussian copula are given as

$$C(u,v|\rho) = C(u,v|\rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \left\{ -\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} \right\},$$

$$c(u,v|\rho) = \frac{1}{(1-\rho^2)^{1/2}} \exp\left\{ \frac{[\rho^2(\Phi^{-1}(u))^2 - 2\rho\Phi^{-1}(u)\Phi^{-1}(v) + \rho^2(\Phi^{-1}(v))^2]}{2(1-\rho^2)} \right\},$$

$$C_{u|v}(u|v;\rho) = \Phi\left(\frac{\Phi^{-1}(u) - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}} \right),$$

$$C_{u|v}^{-1}(u|v;\rho) = \Phi\left(\Phi^{-1}(u)\sqrt{1-\rho^2} + \rho\Phi^{-1}(v) \right).$$

$\Phi^{-1}(.)$ is the standard normal quantile function. $u, v \in (0,1)$, and the correlation parameter $\rho \in (-1,1)$.

A.1.5 Student-t Copula

Student-t copula is an Elliptical copula and is symmetric. It has tail dependency dictated by the degrees of freedom. It is two-parameter based copula. The various transformations for Student-t copula are given as

$$C(u,v|\Theta) = C(u,v|\rho,\nu) =$$
$$\int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{\Gamma(\frac{(\nu+2)}{2})}{\Gamma(\frac{\nu}{2})\pi\nu(1-\rho^2)^{\frac{1}{2}}} \exp\left\{ 1 + \frac{(x^2 - 2\rho xy + y^2)}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2},$$

$$c(u,v|\rho,\nu) = \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\sqrt{1-\rho^2}[\Gamma(\frac{\nu+1}{2}]^2]},$$
$$\times \frac{([1 + \frac{(t_\nu^{-1}(u))^2}{\nu}][1 + \frac{(t_\nu^{-1}(v))^2}{\nu}])^{\frac{\nu+1}{2}}}{[1 + \frac{(t_\nu^{-1}(u))^2 + (t_\nu^{-1}(v))^2 - 2\rho t_\nu^{-1}(u) t_\nu^{-1}(v)}{\nu(1-\rho^2)}]^{\frac{\nu+2}{2}}},$$

$$C_{u|v}(u|v;\rho,\nu) = t_{\nu+1}\left\{ \frac{t_\nu^{-1}(u) - \rho t_\nu^{-1}(v)}{\sqrt{\frac{\nu + (t_\nu^{-1}(v))^2(1-\rho^2)}{\nu+1}}} \right\},$$
$$C_{u|v}^{-1}(u|v;\rho,\nu) = t_\nu\left\{ t_{\nu+1}^{-1}(u)\sqrt{\frac{\nu + (t_\nu^{-1}(v))^2(1-\rho^2)}{\nu+1}} + \rho t_\nu^{-1}(v) \right\},$$

where $\Gamma(a)$ is the gamma function equal to

$\Gamma(a) = \int_0^\infty x^(a-1)e^{-x}dx.$

$t^{-1}(.)$ denotes the standardized student-t quantile function. $u, v \in [0,1]$ and the correlation paramater $\rho \in (-1,1)$, and the degrees-of-freedom parameter $\nu > 0$.

§ A.2 Re-Parametrization

Copula families generally vary across in terms of their respective parameters ranges. We propose a re-parameterization for a copula. Such a method unifies the range and makes the proposal's distribution support over the real line $\mathbb{R}$.

If the transformation is $\mathbf{Z} : \Theta \mapsto \Psi$, then for various copula families,

Clayton Copula

$\alpha \in (0,\infty), \qquad \mathbf{Z}(\alpha) = \exp(\psi).$

Gumbel Copula

$\alpha \in [1,\infty), \qquad \mathbf{Z}(\alpha) = \exp(\psi) + 1.$

Gaussian Copula

$\rho \in [-1,1], \qquad \mathbf{Z}(\rho) = \dfrac{1-e^{-\psi}}{1+e^{-\psi}}.$

Student-t Copula

$\rho \in [-1,1], \qquad \mathbf{Z}_1(\rho) = \dfrac{1-e^{-\psi_1}}{1+e^{-\psi_1}}.$

$\nu \in [1,\infty), \qquad \mathbf{Z}_2(\nu) = \exp(\psi_2) + 1.$

§ A.3 Sampling from a Truncated Copula

Each u_t is sampled from the conditional copula distribution conditional on u_{t-1} and truncated through the order statistics. As we showed before the neighbouring instances are simply to ensure truncation.

Let a and b be the lower and upper limit, for a draw u to lie in. Then as seen, the conditional copula distribution is given as

$$C_{u|v}(u|v;\Theta) = \frac{\partial C(u,v|\Theta)}{\partial v}.$$

Let $C^{Tr}_{u|v}(u|v;\Theta,a,b)$ be the truncated conditional copula, such that

$$\begin{aligned} C^{Tr}_{u|v}(u|v;\Theta,a,b) &= \frac{\int_a^u c_{t|v}(t|v;\Theta)dt}{C_{b|v}(b|v;\Theta) - C_{a|v}(a|v;\Theta)}, \\ &= \frac{C_{u|v}(u|v;\Theta) - C_{a|v}(a|v;\Theta)}{C_{b|v}(b|v;\Theta) - C_{a|v}(a|v;\Theta)}. \end{aligned}$$

Let $C^{Tr}_{u|v}(u|v;\Theta,a,b) = w$, then we can re-arrange equation above to

$$w \cdot [C_{b|v}(b|v;\Theta) - C_{a|v}(a|v;\Theta)] + C_{a|v}(a|v;\Theta) = C_{u|v}(u|v;\Theta).$$

Let the L.H.S in above equation be equal x, so

$x = C_{u|v}(u|v;\Theta)$, where now by simply inverting the conditional distribution we get

$$u = C^{-1}_{x|v}(x|v;\Theta).$$

$u \in (a,b)$. Hence we simply need the inverse of the conditional distribution to successfully draw truncated instances.

Bibliography

Al-Osh, M. A. and Alzaid, A. A. First-order integer-valued autoregressive (inar(1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987. 83

Beare, Brendan K. Copulas and temporal dependence. *Econometrica*, 78(1):395–410, 2010a. 64, 69, 70

Beare, Brendan K. Archimedean copulas and temporal dependence. University of california at san diego, economics working paper series, Department of Economics, UC San Diego, Sep 2010b. 64, 70

Boero, Gianna; Silvapulle, Param, and Tursunalieva, Ainura. Modelling the bivariate dependence structure of exchange rates before and after the introduction of the euro: a semi-parametric approach. *International Journal of Finance & Economics*, 16(4): 357–374, 2011. 7, 8, 15, 29

Bouyé, Eric; Durrleman, Valdo; Nikeghbali, Ashkan; Riboulet, Gaël, and Roncalli, Thierry. Copulas for finance - a reading guide and some applications. *Social Science Research Network Working Paper Series*, November 2007. 2

Bouyé, Eric; Salmon, Mark Howard, and Gaussel, Nicolas. Investing dynamic dependence using copulae. *Working Paper*, 2008. 12

Cameron, A. Colin; Li, Tong; Trivedi, Pravin K., and Zimmer, David M. Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *Econometrics Journal*, 7(2):566–584, December 2004. 2

Carley, Holly. Maximum and minimum extensions of finite subcopula. *Communications in Statistics - Theory and Methods*, 31(12):2151–2166, 2002. doi: 10.1081/STA-120017218. 74

Chen, Xiaohong and Fan, Yanqin. Estimation of copula-based semiparametric time series models. Working Papers 0226, Department of Economics, Vanderbilt University, October 2002. 64, 69, 72, 88, 89

Chen, Xiaohong; Wu, Wei Biao, and Yi, Yanping. Efficient estimation of copula-based semiparametric markov models. Cowles Foundation Discussion Papers 1691, Cowles Foundation for Research in Economics, Yale University, February 2009. 64, 70, 72, 88

Cherubini, U. and Luciano, E. Value at Risk Trade-off and Capital Allocation with Copulas. *Economic Notes*, 30:235–56, 2001. 2

Chib, Siddhartha and Greenberg, Edward. Analysis of multivariate probit models. *Biometrika*, pages 347–361, 1998. 46, 81

Chib, Siddhartha and Winkelmann, Rainer. Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–35, 2001. 2, 46, 81

W.Darswo, B. Nguyen and Olsen, E. Copulas and markov processes. *Illinois Journal of Mathematics*, 36:600–642, 1992. 63, 64, 66, 67, 68

Demarta, Stefano and McNeil, Alexander J. The t copula and related copulas. *International Statistical Review*, 73:111–129, 2005. 2

Dias, Alexandra and Embrechts, Paul. Modeling exchange rate dependence dynamics at different time horizons. *Journal of International Money and Finance*, 29(8):1687–1705, December 2010. 7, 8, 15, 28

Diebold, F. X.; Gunther, T., and Tay, A.S. Evaluating Density Forecasts, with Applications to Financial Risk Management. *International Economic Review*, 39:863–83, 1998. 3

Doukhan, Paul; Fermanian, Jean-David, and Lang, Gabriel. An empirical central limit theorem with applications to copulas under weak dependence. *Statistical Inference for Stochastic Processes*, 12(1):65–87, 2009. 73

Embrechts, P.; McNeil, A., and Straumann, D. Correlation and Dependence in Risk Management: Properties and Pitfalls In: Risk Management: Value at Risk and Beyond. *Cambridge University Press, Cambridge*, pages 176–223, 2002. 1

Embrechts, P.; Lindskog, F., and Mcneil, A. Modelling Dependence with Copulas and Applications to Risk Management. In *Handbook of Heavy Tailed Distributions in Finance*, chapter 8, pages 329–384. 2003. 2

Embrechts, Paul; McNeil, Alexander, and Straumann, Daniel. Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value At Risk And Beyond*, pages 176–223. Cambridge University Press, 1999. 2

Engle, R. F. and Kroner, K. F. Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11:122–50, 1995. 1

Engle, R.F. and Manganelli, S. CAViaR: Conditional autoregressive Value at Risk by Regression Quantile. *Journal of Business and Economic Statistics*, 22:367–381, 2004. 1

Engle, Robert. Dynamic conditional correlation. *Journal of Business and Economic Statistics*, 20(3):339–350, 2002. 1, 17, 18

Fermanian, Jean-David and Scaillet, Olivier. Nonparametric estimation of copulas for time series. Fame research paper series, International Center for Financial Asset Management and Engineering, February 2003. 73

Freeland, R.K. and McCabe, B.P.M. Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3):427 – 434, 2004. 83

Genest, C. and Nešlehová, J. A primer on copulas for count data. *Astin Bulletin*, 37(2): 475, 2007. 38, 43, 47, 73, 74, 75

Genest, C.; Ghoudi, K., and Rivest, L.-P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3): 543–552, 1995. 4, 8, 15, 16, 28, 43, 44, 61

Hansen, B.E. Autoregressive Conditional Density Estimation. *International Economic Review*, 35:705–30, 1994. 1

Harvey, C. R. and Siddique, Akhtar. Autoregressive Conditional Skewness. *The Journal of Financial and Quantitative Analysis*, 34(4):465–487, December 1999. 1

Hoff, Peter D. Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.*, 1(1):265–283, 2007. 2, 4, 5, 39, 43, 47, 48, 50, 61, 64, 65, 75, 81, 89

Ibragimov, Rustam. Copula-based characterizations for higher order markov processes. *Econometric Theory*, 25(03):819–846, June 2009. 64, 89

Joe, H. *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, 1997. 3, 11, 16, 64, 72, 88

Kim, Gunky; Silvapulle, Mervyn J., and Silvapulle, Paramsothy. Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51(6):2836–2850, March 2007. 8, 16, 17, 28

Lentzas, G and Ibragimov, Rustam. Copula and long memory. *Harvard Institute of Economic Research*, (2160), 2008. 64, 68, 88

Longin, F. and Solnik, B. Extreme Correlation of International Equity Markets. *Journal of Finance*, 56(2):649–676, 2001. 7, 12

MacDonald, I.L. and Zucchini, W. *Hidden Markov and Other Models for Discrete- valued Time Series: A Practical Introduction using R*. Monographs on Statistics and Applied Probability. Taylor & Francis, 1997. ISBN 9780412558504. 84

Munkin, Murat K. and Trivedi, Pravin K. Simulated maximum likelihood estimation of multivariate mixed-poisson regression models, with application. *Econometrics Journal*, 2(1):29–48, 1999. 2

Nelsen, R. B. *An Introduction to Copulas*. Springer, 2007. 70

Nelson, R.B. *An Introduction to Copula*. Springer Series in Statistics, Springer, New York, second edition, 2006. 3, 10

Patton, A. J. Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, 47(2):527–556, 2006. 2, 7, 8, 12, 13, 15, 17, 23, 28

Pitt, Michael; Chan, David, and Kohn, Robert. Efficient bayesian inference for gaussian copula regression models. *Biometrika*, 93(3):537–554, September 2006. 38, 39, 42, 46, 47, 48, 61, 64, 75, 81

Rényi, A. On measures of dependence. *Acta Mathematica Hungarica*, 10:441–451, 1959. 70

Rodriguez, Juan Carlos. Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, 14(3):401–423, June 2007. 2

Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l Institut Statistique de l'Univwesitè de Paris*, 8:229–31, 1959. 1, 3, 68

Smith, Michael S. and Khaled, Mohamad A. Estimation of copula models with discrete margins via Bayesian data augmentation. 107(497):290–303, 2012. 64, 75

Smith, Murray D. Modelling sample selection using archimedean copulas. *Econometrics Journal*, 6(1):99–123, 2003. 2

Song, Peter Xue-Kun. Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320, 2000. 41

Takagi, S. The Yen and its East Asian Neighbors, 1980-95: Cooperation or Competition? *In T. Ito and A. O. Krueger, eds., Changes in Exchange Rates in Rapidly Developing Countries*, pages 185–207, 1999. 6

Trivedi, Pravin K. and Zimmer, David M. Copula Modeling: An Introduction for Practitioners. *Foundations and TrendsÆ in Econometrics*, 1(1):1–111, 2006. ISSN 1551-3076. 2, 38, 50, 75

Volkonskii, VA and Rozanov, Yu A. Some limit theorems for random functions. i. *Theory of Probability & Its Applications*, 4(2):178–197, 1959. 70

Yu, Bin. Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability*, pages 94–116, 1994. 70

Zimmer, David M. and Trivedi, Pravin K. Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business & Economic Statistics*, 24:63–76, 2006. 2